Cable security state detection based on multi-scale information fusion and distributed fiber optic sensing

Senqing Huang[3*], Qijiu Xia[3], Zhenyu Qi[3], Kai Qi[1], Wenjun Lv[1] and Yunbo Zhao[1,2]

[1]Department of Automation, University of Science and Technology of China, Hefei, 230026, Anhui, China
[2]Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, 230088, Anhui, China
[3]Institute of Advanced Technology, University of Science and Technology of China, Hefei, 230031, Anhui, China

[*]Corresponding author's e-mail: sa22218208@mail.ustc.edu.cn

Abstract. Mechanical operations in proximity to cables present considerable risks to the stability of power transmission and the reliability of energy supply systems. As an advanced intelligent sensing technique, Distributed Fiber Optic Sensing (DFOS) has the capability to detect and identify potential external hazards affecting cables. However, due to the massive volume of sensing data, existing studies primarily rely on classification algorithms to distinguish threat events, lacking efficient end-to-end detection algorithms for differentiation and localization. This constraint results in inefficient information usage and limited real-time responsiveness. To overcome these challenges, this study introduces a high-performance detection algorithm that integrates multi-scale information fusion. First, an attention mechanism called MRFA is proposed to achieve effective feature extraction, which is characterized by its flexibility and multi-receptive fields. Second, an innovative Information Dialysis Module (DM) is proposed to enhance the efficiency of inter-layer information filtering in detection models. Finally, the proposed methods are integrated into an improved YOLOv8 framework. Experimental comparisons across multiple datasets validate the proposed method's effectiveness and efficiency, demonstrating its capability for real-time surveillance and smart recognition in cable security applications.

1. Introduction

Electricity is vital for modern industry and national energy supply. In special areas like towns, farmlands, and nature reserves, transmission cables are often buried underground. Damage to these cables can lead to severe safety incidents, including economic losses, casualties, and environmental harm. Traditional warning methods, such as stakes and signs, are costly and fail to provide early warnings. Real-time monitoring of intrusion events is essential to issue timely alerts and prevent potential threats.

Distributed optical fiber sensing (DOFS) is an advanced technique that leverages the backscattered Rayleigh signal within a single optical fiber to detect and measure environmental physical parameters along its length. Additionally, underground cables are often buried alongside communication optical fibers, enabling the monitoring of threat events near the cables without requiring additional fiber installation. With benefits such as long sensing distances, anti-interference capabilities, and low installation costs, distributed fiber optic sensing systems have emerged as promising solutions for geophysics and linear infrastructure monitoring [1].

DOFS can collect rich data, which provides a solid foundation for AI modeling. Zhu et al. [2] introduced a pipeline radial threat recognition model that integrates multidimensional information fusion with a broad learning system (MIFBLS) to improve energy pipeline security. The study addresses limitations in real-time performance and information utilization in distributed acoustic sensing (DAS) systems, leading to improved signal processing, feature extraction, and incremental learning. He et al. [3] developed a two-stage recognition network and carried out field experiments to evaluate its accuracy in identifying five types of intrusion events, including animal and human intrusions as well as mechanical excavation, achieving an average recognition rate of 97.04%. In addition, Yang et al. [4] introduced a semi-supervised learning approach for remote pipeline intrusion monitoring, significantly enhancing the detection and localization of intrusion events even under low signal-to-noise ratio conditions.

However, most of the above research methods rely on image classification algorithms for detection, making the detection performance dependent on segmenting high-quality raw signal slices and unable to achieve classification and localization within the algorithm. Therefore, it is necessary to employ object detection methods to identify the types and locations of threat events from large volumes of sensing data. However, object detection algorithms need to be adapted to accommodate the massive data characteristics of fiber optic sensing as much as possible.

To fulfill the high-efficiency detection demands in Distributed Fiber Optic Sensing (DFOS) applications, this paper presents a target detection model incorporating multi-scale information fusion. Initially, a plug-and-play Multi-Receptive Field Attention mechanism (MRFA) is introduced to enhance feature extraction efficiency within the object detection framework. Next, we propose an innovative Information Dialysis Module (DM), which adopts a novel high-efficiency filtering framework to improve parameter efficiency. Finally, the proposed methods are integrated into an improved YOLOv8 framework. Comparative evaluations on public and real-world cable monitoring datasets show that the proposed method improves mAP by 3%-5% while reducing the number of parameters by nearly 50% compared to models with comparable performance.

The key contributions of this study are as follows:

1. To enhance feature extraction efficiency in object detection models, we introduce a Multi-Receptive Field Attention mechanism (MRFA), which leverages depth-wise separable convolutions for cross-spatial learning. It improves feature extraction efficiency with minimal parameter overhead.
2. To optimize baseline network architectures, we propose an Information Dialysis Module (DM) that incorporates design principles such as hierarchical feature aggregation and cross-stage partial connections. This architectural improvement enhances the utilization efficiency of learnable parameters.
3. To integrate our proposed methods into YOLO, we present an innovative object detection framework that combines the DM module and MRFA attention mechanism. In object detection tasks, this framework significantly enhances the mean average precision (mAP) of detection.

2. Related work

Related work will be introduced from three aspects: Object Detection, Architectures Design, and Attention Mechanism.

2.1. Object detection

Object detection, which is a fundamental task in computer vision, focuses on detecting and pinpointing the locations of objects within images. Convolutional neural networks (CNNs) have emerged as a primary tool for real-time detection, offering a good trade-off between performance and efficiency, with the YOLO (You Only Look Once) models being particularly prominent[5]. Introduced in 2016, YOLOv1 pioneered single-stage detection with anchor boxes, and YOLOv2 improved localization accuracy. YOLOv3 advanced the field with novel loss functions like CIoU and GIoU and multi-scale feature extraction. YOLOv4 introduced the CSPDarknet53 framework, enhancing both accuracy and speed. YOLOv5 refined the CSP approach, achieving an impressive 200 FPS for real-time applications. Subsequent versions, YOLOv6 and YOLOv7, further optimized detection accuracy and inference efficiency, with YOLOv7 excelling in speed and accuracy through composite scaling methods. The latest, YOLOv8, introduced anchor-free models, maintaining high precision with a remarkable 280 FPS. New efficient modules are constantly being proposed in the YOLO algorithm, but there is still a lack of further innovation in its architecture. This paper aims to make innovations in two aspects: enhancing feature extraction through the introduction of an efficient attention mechanism, and improving feature fusion with the proposal of an Information Dialysis Module.

2.2. Architectures design
Efficient and high-quality network architecture design has always been a goal for researchers. In the area of designing convolutional neural network (CNN) architectures, ResNeXt first demonstrated that cardinality is more effective than width and depth dimensions. DenseNet links the output features of all previous layers to serve as input for the subsequent layer, a technique that can be viewed as maximizing cardinality. CSPNet proposed cross-stage partial networks to optimize redundant gradient information. It enhances the diversity of gradients by combining feature maps from both the early and late stages of the network. Gold-YOLO [6] suggested that feature pyramid networks (FPN) and path aggregation networks (PANet) merely mitigate the issue of information fusion and offer a more sophisticated gather-and-distribute (GD) mechanism, which improves multi-scale feature fusion through convolutional and self-attention operations. However, the work of Gold-YOLO is primarily applied to neck feature fusion architectures. The issue of feature fusion still exists in the earlier backbone networks. These architectural design works inspire us to believe that the key to improving feature fusion capability lies in aggregating and reintegrating multi-level feature information.

2.3. Attention mechanism
The attention mechanism improves feature extraction, and it typically includes three primary types: channel attention, spatial attention, and a combination of both channel and spatial attention. SE (Squeeze-and-Excitation) emphasizes the role of inter-dimensional interactions by assigning different attention weights to different channels. The Convolutional Block Attention Module (CBAM) builds cross-channel and cross-spatial relationships by leveraging the semantic dependencies between the spatial and channel dimensions of feature maps, showing the ability to improve features by incorporating cross-dimensional attention weights into the input. The Efficient Multi-scale Attention (EMA) [7] introduces an innovative multi-scale attention module that minimizes computational costs by restructuring certain channels into the batch dimension and organizing the channel dimensions into several sub-features, ensuring a balanced distribution of spatial semantic features across each feature group. However, the cross-spatial learning strategy in EMA cannot fuse larger receptive field spatial information into the channel dimension. The MRFA (Multi-Receptive Field Attention) proposed in this paper efficiently integrates spatial information from multiple receptive fields into the channel dimension through an optimized architecture.

3. Method
This section offers a summary of the MRFA, DM module, and the enhanced YOLO architecture.

3.1. Multi-receptive field attention (MRFA)
The MRFA attention module incorporates 3×3, 5×5, and 7×7 depthwise separable convolution kernels within the 3×3 branch to capture multi-scale feature representations across different layers. Additionally, feature maps of different depths are transmitted to the next stage through three distinct information flows. Consequently, MRFA not only captures information across channels to modulate the significance of different channels, but also maintains multi-scale spatial structure and semantic information within the channels. As shown in Figure 1, this illustrates the structure of the proposed MRFA.

When a feature map passes through the MRFA, the first step is grouping. During the grouping process, the feature map is divided along the channel dimension into different groups, and the feature maps from different groups are processed in parallel, improving computational efficiency. Then, within each group, the feature map undergoes four distinct branches. One branch remains unchanged, while the other three branches participate in extracting weights for reconstruction across different receptive fields and dimensions. Finally, the obtained weights are multiplied with the original feature map to produce the adjusted feature map with the final weights.

It is important to note that among the three branches involved in weight reconstruction, the X-average pooling branch and the Y-average pooling branch belong to the same category of weight adjustment methods, while the rightmost multi-receptive field branch belongs to another specialized information modulation method. The weight adjustment along the X or Y directions is mainly achieved by reducing the dimensionality of the feature map in the spatial direction along a certain axis to extract the weights, and its characteristic is that it is flattened. In contrast, the multi-receptive field branch we emphasize achieves a three-dimensional weight extraction through the extraction of Multi-layered spatial information. It is described as three-dimensional because, before performing the flattening dimensionality reduction, it uses efficient depthwise separable convolutions stacked into a feature pyramid-like structure. This enables it to capture information at different scales, which is then incorporated into the final weight reconstruction process.

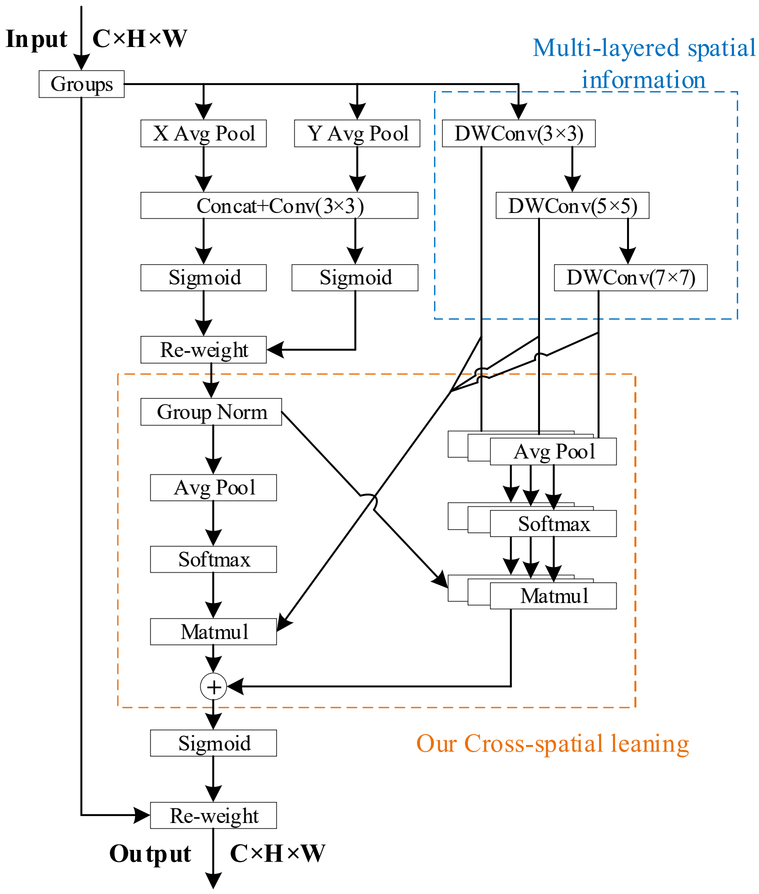

Figure 1. The architecture of the proposed MRFA.

3.2. Design of the DM module

We compare the evolution of the Bottleneck structure from YOLOv3 to YOLOv5 and then to YOLOv8. First, the channel compression ratio has been adjusted to 1, and the initial 1×1 standard convolution has been replaced by a 3×3 convolution layer. It is evident that the Bottleneck-like structure increasingly resembles a multi-layer feature pyramid structure, with the only difference being that the feature map scale does not decrease as the number of layers increases. Inspired by this, we designed an internal module in the channel dimension to enable information interaction and filtering along this dimension, as shown in Figure 2.

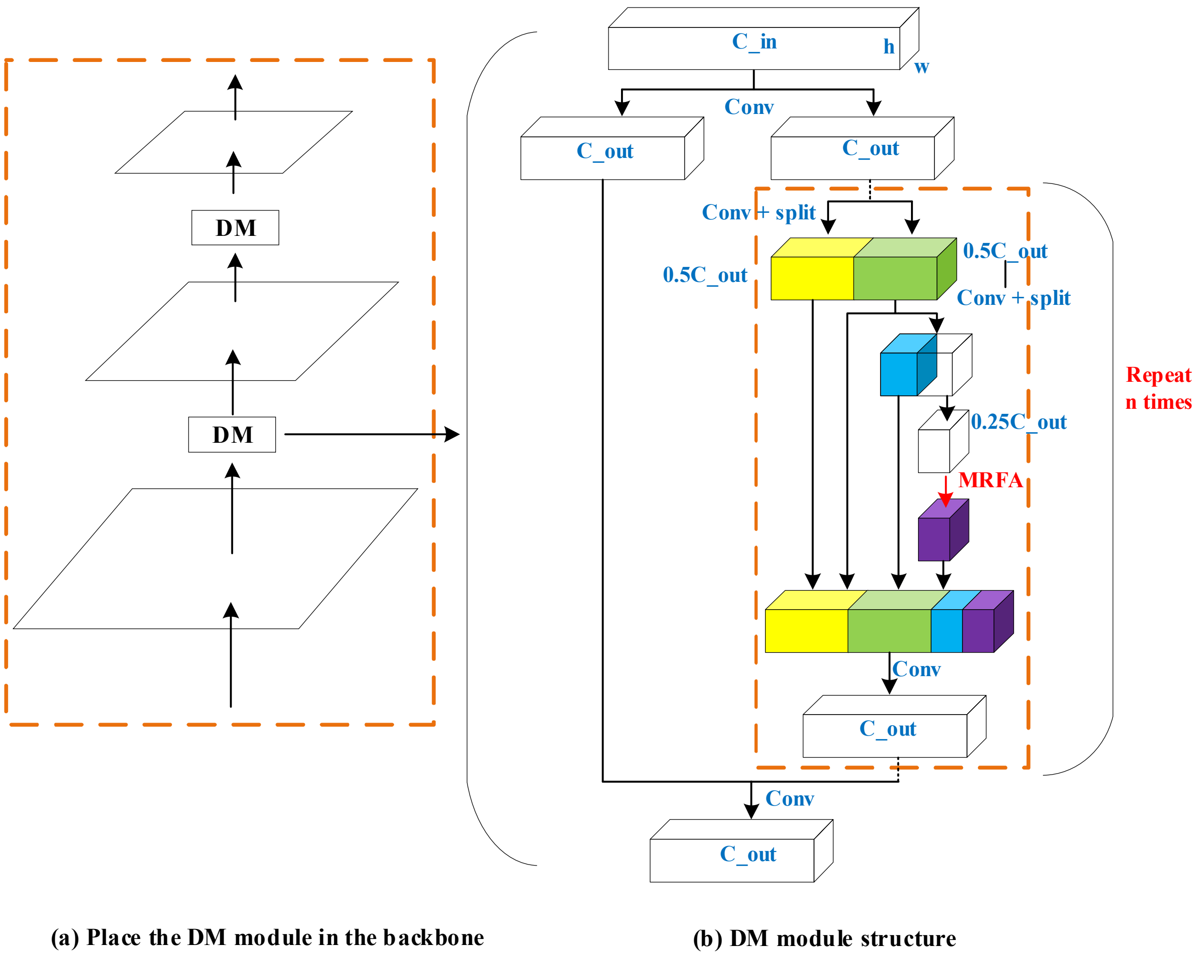

Figure 2. DM module. (a) Placement of the DM module in the backbone network; (b) Detailed structural design of the DM module.

The developers of YOLOv7 introduced an extension of ELAN to improve the network's learning capacity while maintaining the original gradient pathways. In the design of our DM module, we aimed to adhere to the principles of excellent architectures like ELAN. Specifically, our strategy includes using multiple segmentation operations to divide information flows, inspired by the CSPNet concept. Additionally, we designed a three-layer feature pyramid structure to provide a semantic summary of the current input feature map. Each layer partitions a portion of the information flow and directly incorporates it into the final integration stage. This not only retains the semantic information of each layer but also further increases the channel base.

When a feature map passes through the DM module, it first undergoes a standard convolution module. After this, the input channel number is adjusted to match the output channel number. The input and output channel numbers serve as variable parameters for the DM module, allowing it to connect two modules with either the same or different channel numbers. This flexibility enables the DM module to be seamlessly inserted into any structure of a detection network. Next, after passing through the first standard convolution module, the feature map splits into two branches. One branch remains unchanged, while the other participates in a more complex feature fusion process. These two branches are stacked into a feature map with twice the output channel number, and after passing through the final standard convolution module, the channel count is halved, producing the final output of the DM module.

It is important to note that this more complex feature fusion process has a configurable repetition parameter, n, which can be selected according to different requirements. In this feature fusion module, the input feature map undergoes a split operation after each standard convolution block. The split operation divides half of the output channels and only allows half of the channels to flow into the next layer. This helps control the scale of the feature map entering deeper layers, preventing excessive computational resource consumption, while still preserving the original features at each layer for the final output selection. When the number of channels in the feature map is reduced to one-quarter of the original, we use an MRFA module to perform lightweight feature fusion, which is crucial for improving the performance of the network model.

3.3. Scheme for inserting the DM module

The Dialysis theory highlights that the absence of efficient filtering mechanisms between layers limits the performance of multi-layer neural networks. The DM module is thus positioned before information splitting and after aggregation, preserving the original algorithm structure to prevent new chaotic flows. In the YOLOv8n-DM detector, the Backbone-Neck-Head paradigm is followed, retaining the Head for final detection while reconstructing the Backbone and Neck by integrating the DM module.

As illustrated in Figure 3, the DM module is placed after each change in feature map scale within the Backbone to improve information filtering across various semantic levels. In the Neck, the DM module is placed after feature fusion to filter simultaneous information flows from adjacent semantic levels.

It should be noted that there are two types of DM modules within the entire object detection network structure. One type is positioned after standard convolution blocks, primarily located in the Backbone. A notable characteristic of this type is that the input channel number is smaller than the output channel number. The other type is placed after concat modules, where it can normalize the output channel number. This type mainly focuses on summarizing and fusing the information after concatenation.

4. Experimental results and analyses

This section will concentrate on three key areas: evaluation metrics, datasets, and the specific results of comparative experiments along with their analysis.

4.1. Model performance evaluation metrics

In object detection, model performance is commonly assessed using the mean Average Precision (mAP), which is determined by the area under the Precision (P) and Recall (R) curves. Precision (P) is defined as:

$$P = \frac{TP}{TP + FP} \tag{1}$$

where TP denotes true positives and FP denotes false positives. Precision measures the likelihood of the model correctly identifying positive instances. Recall (R) is defined as:

$$R = \frac{TP}{TP + FN} \tag{2}$$

where FN denotes false negatives. Recall indicates how many objects the model correctly identifies. When calculating mAP, the size of the Intersection over Union (IoU) is used as a criterion. The closer the IoU is to 1, the more accurate the detection box is. For each different IoU threshold, mAP can be calculated for different categories. When the mean Average Precision is calculated at an Intersection over Union (IoU) threshold of 0.5, the resulting value is referred to as $mAP@50$. The specific definition is as follows:

$$mAP@50 = \frac{\sum_{i=1}^{N} AP(i)}{N} \tag{3}$$

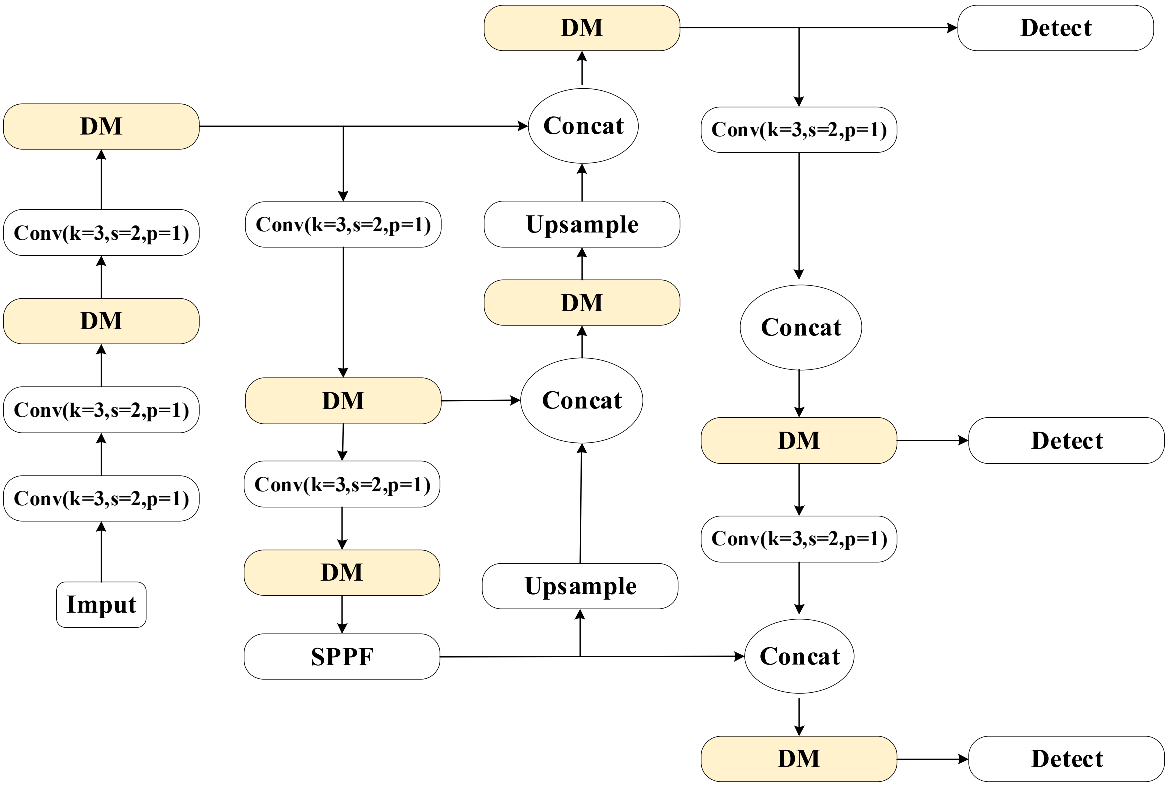

Figure 3. The figure illustrates the overall design scheme for inserting the DM module into the YOLO series v8n algorithm.

where N denotes the number of classes, and $AP(i)$ denotes the Average Precision (AP) for class i. AP represents the accuracy of a single class, and its calculation involves generating Precision-Recall curves and computing the area under the curve. $mAP@50:95$ represents the average precision over the range of Intersection over Union (IoU) thresholds from 0.5 to 0.95. It offers a thorough evaluation of the model's performance at various IoU thresholds. The specific definition is as follows:

$$mAP@50:95 = \frac{\sum_{t=0.5}^{0.95} \sum_{i=1}^{N} AP(i)}{10N} \tag{4}$$

4.2. Dataset

Firstly, we evaluate our method using two public datasets: NEU-DET for industrial defect detection and PASCAL VOC2012 for general computer vision tasks. Additionally, we collect monitoring data from real-world wind power underground cable scenarios, covering various safety events like excavator operation, vehicle movement, and manual excavation. Around 1600 processed raw signals are used for training the detection network. Specifically, the distributed fiber optic sensor data acquisition card collects tens of thousands of sampling points on the entire optical fiber at a rate of 2000HZ, which can eventually be processed into a spatiotemporal graph with the horizontal axis representing the spatial direction and the vertical axis representing the time direction. Different events correspond to different spatiotemporal graphs. The labeled dataset is put into the target detection model for training, and the obtained model parameters are used for inference of unlabeled data.

4.3. Results

The experimental results focus on three aspects: the effectiveness of the DM module, the MRFA ablation study, and detection performance on real distributed fiber optic sensing data. Table 1 shows that our method improves performance by 3% to 5% while reducing the parameter size by about 50% compared to similar models.

Table 1. Performance comparison of baseline models with and without the DM module.

Models	Parameters/M	Cost/GFLOPs	VOC		NEU-DET	
			mAP@50	mAP@50:95	mAP@50	mAP@50:95
V8n	3.16	8.9	0.6473	0.4635	0.743	0.400
V8n-DM(our)	**6.30**	**15.8**	**0.6875**	**0.4984**	**0.763**	**0.416**
V8s	11.17	28.8	0.7001	0.5092	0.746	0.411
V3n	3.88	11.2	0.6361	0.4457	0.746	0.394
V3-DM(our)	**6.42**	**18.7**	**0.6660**	**0.4730**	**0.735**	**0.399**
V3s	14.60	40.3	0.6962	0.5042	0.739	0.405
V5n	2.51	7.1	0.6280	0.4340	0.728	0.394
V5n-DM(our)	**6.28**	**17.0**	**0.6840**	**0.4940**	**0.746**	**0.408**
V5s	9.13	24.1	0.6939	0.4910	0.718	0.401
V6n	4.24	11.9	0.6387	0.4639	0.752	0.408
V6n-DM(our)	**6.66**	**19.2**	**0.6720**	**0.4900**	**0.757**	**0.413**
V6s	16.31	44.2	0.6929	0.5110	0.746	0.416

Table 2 shows that incorporating MRFA improves performance by 1% to 2% with minimal additional parameters and computation.

Table 2. Performance comparison of models with and without the MRFA module.

Models	MRFA	Parameters/M	Cost/GFLOPs	VOC		NEU-DET	
				mAP@50	mAP@50:95	mAP@50	mAP@50:95
V8n-DM	√	**6.07**	**16.0**	**0.688**	**0.498**	**0.763**	**0.416**
		6.06	18.8	0.679	0.492	0.751	0.395
V3n-DM	√	**6.78**	**20.6**	**0.666**	**0.473**	0.735	**0.399**
		6.77	20.3	0.663	0.472	**0.755**	0.375
V5n-DM	√	**6.27**	**16.9**	**0.684**	0.494	**0.746**	**0.408**
		6.27	16.7	0.679	**0.497**	0.735	0.407
V6n-DM	√	**6.91**	**19.5**	**0.672**	**0.490**	**0.757**	**0.413**
		6.91	19.2	0.667	0.485	0.756	0.413

Table 3 displays the performance of our proposed method on an actual distributed fiber optic sensing dataset, showing that the DM module consistently strikes a better balance between efficiency and accuracy.

Table 3. Evaluation of the proposed method's performance on the DFOS dataset.

Models	Parameters/M	Cost/GFLOPs	mAP@50	mAP@50:95
V8n	3.16	8.9	0.772	0.498
V8n-DM(our)	**6.30**	**15.8**	**0.807**	**0.512**
V8s	11.17	28.8	0.811	0.514

5. Conclusion
This paper develops an object detection architecture applicable to distributed fiber optic sensing data, specifically incorporating the Multi-Receptive Field Attention mechanism (MRFA) and the Information Dialysis Module (DM). Experimental results demonstrate that this architecture achieves a 3% to 5% performance improvement while maintaining only about 50% of the parameter size of models with similar accuracy. Moreover, the proposed method can be seamlessly adapted to different computer vision tasks to deliver optimal performance. We hope the proposed structure provides valuable insights for CNN architecture design.

Acknowledgments
This research was partially funded by the National Key Research and Development Program of China under Grants 2022YFB3206900 and 2022YFB3206904.

References
[1] Ma, Z., Zhang, M. J., Jiang, J. F., Zhang, J. Z., Xiao, L. T., & Liu, T. G. (2023). Fiber distributed acoustic sensing technology based on linear frequency modulation pulses. Laser & Optoelectronics Progress, (11), 24-43. (in Chinese).
[2] Zhu, C., Pu, Y., Lyu, Z., Yang, K., & Yang, Q. (2024). Multidimensional information fusion and broad learning system-based condition recognition for energy pipeline safety. Knowledge-Based Systems, 300, 112259.
[3] He, T., Liu, Y., Zhang, S., Yan, Z., Liu, D., & Sun, Q. (2021, May). High accuracy intrusion pattern recognition using a dual-stage-recognition network for fiber optic distributed sensing system. In CLEO: QELS_Fundamental Science (pp. JW1A-119). Optica Publishing Group.
[4] Yang, Y., Zhang, H., & Li, Y. (2021). Long-distance pipeline safety early warning: a distributed optical fiber sensing semi-supervised learning method. IEEE Sensors Journal, 21(17), 19453-19461.
[5] Hussain, M. (2024). Yolov1 to v8: Unveiling each variant–a comprehensive review of yolo. IEEE Access, 12, 42816-42833.
[6] Wang, C., He, W., Nie, Y., Guo, J., Liu, C., Wang, Y., & Han, K. (2024). Gold-YOLO: Efficient object detector via gather-and-distribute mechanism. Advances in Neural Information Processing Systems, 36.
[7] Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., & Huang, Z. (2023, June). Efficient multi-scale attention module with cross-spatial learning. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.