

Multi-scale fire target detection algorithm using YOLO-fire

FAN Weiqiang^{1,2*}, DING Jiayan¹, PENG Bin¹, LIU Dong¹, GAO Shuoheng¹, JIA Changzhuo¹

1. School of Electronic Information Engineering, Inner Mongolia University, Hohhot 010021, China;

2. Inner Mongolia Key Laboratory of Intelligent Communication and Sensing and Signal Processing, Hohhot 010021, China

*Corresponding author: FAN Weiqiang (fan_weiqiang@163.com)

Received: October 11, 2025

Revised: November 17, 2025

Accepted: November 21, 2025

Abstract: Accurate and real-time fire detection is crucial for industrial production and daily life. However, the variable form of fire and the significant differences in visual characteristics across its different stages pose great challenges to precise fire prevention and control. To address this issue, a multi-scale fire target detection algorithm using YOLO-fire was proposed by improving the YOLOv8 model. This model introduced new layer structures and attention mechanism, replaced new feature fusion modules and loss functions. By introducing a small-target detection P2 layer, the model's ability to detect early-stage fires is improved. The coordinate attention mechanism is integrated into the layer structures of multi-scale target detection, enhancing the capture of target location information and channel relationships, thereby focusing more on the target regions. The Neck network structure was optimized by adopting a BiFPN_F strategy for different feature layers, which strengthened the cross-scale representation of fire features and controlled the parameter count of the designed model. The WIoU loss function was employed to optimize the regression process, improving fire source localization accuracy in complex scenarios, enhancing model robustness, and increasing detection precision. Experimental results on fire datasets demonstrated that YOLO-fire could effectively detect multi-scale fire targets in various scenarios. Compared to the baseline model (YOLOv8n), YOLO-fire achieves improvements of 1.37% in accuracy, 1.25% in mAP50-95, and 0.35% in F1-score, while reducing parameters by 3.79%. Furthermore, compared to current mainstream target detection algorithms, YOLO-Fire achieved optimal detection performance while reducing network parameters and computational complexity. This research provided effective technical support for fire safety prevention and control in related fields.

Key words: fire detection; early-stage fire; feature fusion; attention mechanism; loss function; network structure; YOLOv8

0 Introduction

For a long time, frequent fire accidents not only posed serious threats to human life and property safety, but also triggered social panic, disrupted social order and economic development^[1]. Particularly in confined spaces such as indoor venues, tunnels, and mines, if fires cannot be detected in a timely manner and handled effectively, it may lead to severe casualties and substantial property losses^[2,3]. Therefore, conducting research on multi-scale fire target monitoring technology, enabling early detection of fire sources, and facilitating effective emergency response, holds significant importance for protecting lives and property, maintaining social stability, and promoting technological progress.

Traditional fire monitoring methods primarily utilize various types of sensors, such as light, smoke, and temperature sensors, to detect physical parameters like

brightness, smoke, thermal radiation, infrared radiation, and ultraviolet radiation generated by fires near their installation points. An alarm is triggered when the detected parameters exceed a set threshold. However, point-type sensors have limited monitoring coverage, making it difficult to achieve comprehensive surveillance in complex or large spaces. Additionally, non-fire factors can easily cause false alarms, demanding higher sensor accuracy^[4]. Furthermore, with technological advancements, new types of fires, such as those involving lithium batteries and chemical fires, differ from traditional fires, making multi-sensor detection technology inadequate for monitoring diverse fire types. With the development of machine learning and imaging technology, researchers have proposed fire detection methods based on visual features. These methods extract static and dynamic characteristics of flames or smoke from videos and utilize machine learning models to achieve fire detection^[5,6]. However, fire

detection approaches relying on manually designed features are inherently subjective and suffer from limitations such as limited generalization ability, sensitivity to noise, and difficulty adapting to complex scenarios.

With the advancement of deep learning technology, deep learning-based object detection algorithms can automatically learn features from data, addressing the challenges associated with manually designed features in traditional fire detection algorithms while overcoming the issue of poor generalization in detection models. Currently, deep learning-based object detection algorithms are primarily divided into two categories: 1) two-stage algorithm using candidate regions^[7-12], such as R-CNN and faster R-CNN^[7,9], and 2) single-stage algorithm using regression analysis, such as SSD and the YOLO series^[13-16]. Compared to two-stage algorithms represented by R-CNN, one-stage algorithms like YOLO eliminate the candidate box generation stage, overcoming the slow detection speed of two-stage algorithms. Additionally, YOLO algorithms trained on large datasets are suitable for multi-scale object detection, reducing both false positives and false negatives in complex scenarios. However, existing YOLO algorithms still face challenges in fire detection. For instance, when fire or smoke targets in images are large, they tend to blend with the background, while long-distance, small-target fire sources and closely positioned or overlapping fire targets are poorly detected. Therefore, considering the high real-time requirements of fire prevention and control scenarios, we proposed a fire object detection algorithm based on YOLO-fire to improve the response speed and accuracy of fire detection, further promoted the intelligent development of fire prevention and control technology.

The main contributions of this paper are as follows.

1) Based on the baseline model (YOLOv8) and considering the characteristics of fires at different ignition stages, a novel multi-scale fire detection algorithm named YOLO-fire was proposed by introducing new layer structure and attention mechanism, designing a bidirectional feature pyramid network fusion (BiFPN_F) strategy, and replacing the loss function.

2) A small-target detection layer P2 was incorporated into the YOLOv8 network structure to enhance the model's capability in detecting early-stage fires. Simultaneously, a coordinate attention (CA) module was introduced into the neck network to improve the capture of target location information and channel

relationships, thereby enabling better focus on target regions.

3) A novel bidirectional feature pyramid network (BiFPN) was adopted to replace the original path aggregation feature pyramid network (PAFPN) in YOLOv8. A BiFPN_F strategy for target features in adjacent layers was designed to enhance the cross-scale representation of early fire features while reducing the number of model parameters.

4) The WIoU loss function was employed to replace the CIoU loss function, mitigating the neglect of both simple and hard samples and enabling better focus on diverse regression samples. This allowed YOLO-Fire to be effectively applied to multi-scale fire detection in challenging environments such as mines and tunnels.

1 Related work

1.1 Two-stage fire detection algorithms

Two-stage algorithms, represented by convolutional neural networks (CNN), utilize selective search algorithms to generate candidate regions from input images, extract their feature maps, and perform target classification and bounding box regression. This type of fire detection algorithm has poor real-time performance, but it achieves high detection accuracy. Numerous related studies have been reported. For instance, Girshick et al.^[7] applied R-CNN to object detection, significantly improving its accuracy. To further simplify the model, Girshick^[8] proposed fast R-CNN, an improved algorithm using region extraction and RoI pooling technology. He et al.^[9] introduced faster R-CNN, which incorporated a region proposal network (RPN), further enhancing object detection accuracy. Yu et al.^[10] proposed a flame recognition algorithm based on an improved mask R-CNN, introducing a bottom-up feature fusion within the feature pyramid to strengthen the model's ability to detect flame contours. Zhang et al.^[11] utilized CNN to establish a real-time fire detection system for chemical industrial parks. They employed the CNN-YOLOv5 algorithm to train on both chemical industrial park fire datasets and general fire datasets, verifying the feasibility of using CNN for fire detection in chemical industrial parks. Barmoutis et al.^[12] used the faster R-CNN network to generate candidate fire areas, projected these candidate areas into the Grassmannian space to generate point clouds on the manifold, and employed a vector representation method based on local manifold criteria to aggregate the point clouds, thereby achieving flame detection. However, current two-stage

algorithms generally suffer from long processing times, resulting in poor real-time performance during fire detection processes.

1.2 One-stage fire detection algorithms

Compared to two-stage algorithms, one-stage algorithms offer the advantage of strong real-time performance and have gradually been applied to fire detection in various scenarios. Redmon et al.^[13] first proposed the one-stage object detection algorithm YOLO in 2016 to improve the speed of object detection. In 2017, they further introduced the YOLOv2 algorithm^[14], which enhanced detection speed and accuracy using the K-means clustering algorithm, though it still had limited capability in multi-scale object detection tasks. To address the low accuracy of the YOLO algorithm in multi-scale object detection, Liu et al.^[15] integrated the design philosophy of the YOLO model and proposed the SSD algorithm, a one-stage object detection method based on multi-scale features. Building on related improvements, Megvii Technology^[16] introduced the YOLOX model, which significantly improved detection efficiency through techniques such as MixUp data augmentation and the SiLU activation function.

Based on various YOLO algorithms, researchers have continuously improved components such as the backbone network, feature fusion network, and loss function of YOLO model, applying these enhancements to fire detection across different scenarios. For example, Qian et al.^[17] proposed a fire monitoring and alarm system based on a channel-pruned YOLOv3 model, which improved the real-time performance of fire detection by reducing the number of parameters in the YOLOv3 model. Zhao et al.^[18] proposed a small-target detection algorithm named fire-YOLO to address the detection of minuscule targets, fire-like objects, and smoke-like objects in forest fire images. Cheng et al.^[19] combined an improved Gaussian model with the YOLOv2 network for smoke detection, reducing interference from background variations on detection results. Fan^[20] used the lightweight MobileNetv3 as the backbone network for YOLOv4, significantly reducing the model's parameter count while improving detection speed without compromising accuracy. Akhmedov et al.^[21] proposed a ship fire detection model combining a dehazing algorithm with YOLOv10, which enhanced the fire detection capability of the YOLO model under hazy environments. He et al.^[22] utilized the SSD algorithm to construct a network structure suitable for

forest fires and employed the Imgaug data augmentation method to expand the forest fire dataset, achieving flame recognition in forest fire images. He et al.^[23] improved the YOLOv5 model and proposed an efficient dual-channel bottleneck structure YOLO detection algorithm for fire detection, enhancing the ability to detect flames and smoke with irregular shapes.

These technological studies have further reduced the false detection rate in fire detection across various scenarios, and improved both detection accuracy and real-time performance. However, the precision and real-time performance of YOLO algorithms in early-stage and multi-scale fire detection still require further enhancement.

2 Fire detection algorithm using YOLO-fire

2.1 Network structure of YOLOv8 model

The YOLOv8 model primarily consists of three components: the backbone network, the neck network, and the head^[24]. Compared to previous generations of YOLO models, YOLOv8 achieves further performance improvements through the design of a new network structure and training strategies^[25]. The network structure of the YOLOv8 model is illustrated in Fig.1.

The backbone network is responsible for extracting multi-level features from the input images. This part adopts the Darknet-53 architecture, which utilizes an efficient convolutional layer structure to enable the model to perceive shallow features such as edges, textures, and shapes in the images, as well as high-level semantic information. Additionally, the introduction of residual connections in the C2f module not only mitigates the vanishing gradient problem in deep networks but also improves computational efficiency and reduces the number of parameters.

The neck network is responsible for integrating the multi-level features extracted by the backbone network. This part employs the feature pyramid network (FPN) and path aggregation network (PAN)^[24] to achieve bidirectional cross-scale connections, enabling the fusion of semantic information and detailed features, thereby enhancing the detection capability for targets of different scales.

The detection head is responsible for predicting the location and category of targets. This part adopts an anchor-free mechanism, which directly predicts the target center point and bounding box, eliminating the reliance on predefined anchor boxes and reducing the complexity of hyperparameter adjustment.

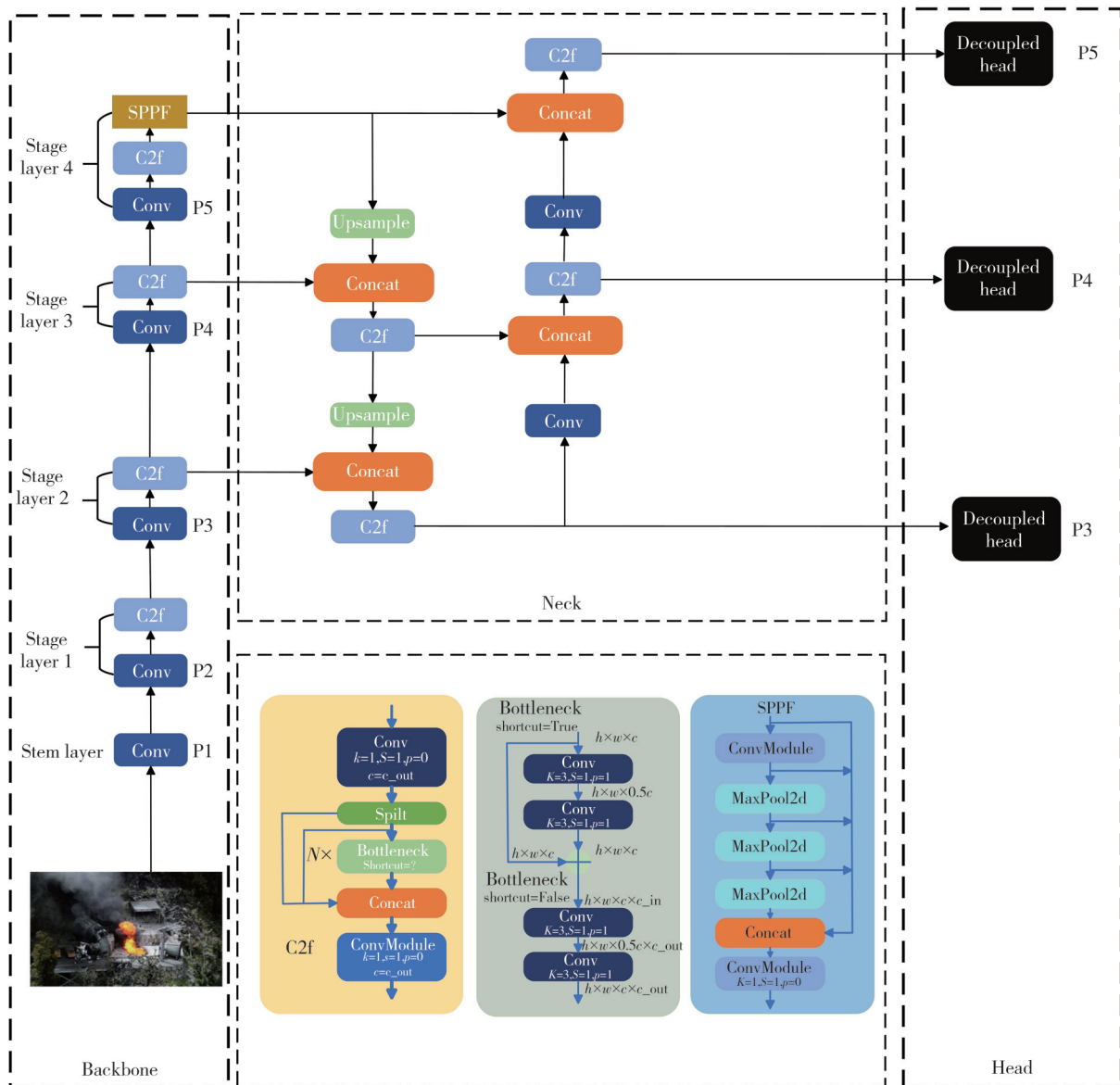


Fig. 1 Network structure of YOLOv8 model

2.2 Network structure of YOLO-fire model

In fire target detection, in order to address the challenges of low detection accuracy for fires and difficulties in detecting multi-scale fire targets in complex environments, we proposed a novel network structure named YOLO-Fire. Based on the YOLOv8 model, this network incorporates four key structural designs and optimizations, aiming to enhance fire detection accuracy at various stages without increasing the number of model parameters. The network structure of the YOLO-Fire model is illustrated in Fig.2.

2.2.1 P2 detection head

The original YOLOv8 structure employs three detection heads (P3, P4, P5), which represent the model's detection capabilities for smaller, medium, and large-scale targets, respectively. While this design

features a low parameter count, it has limited perception of small target information, resulting in poor detection performance for small objects. Fire sources in the early stages of a fire are typical small targets. To adapt to fire source detection across different fire stages, this paper adds a P2 detection head to the YOLOv8 model. By enhancing information transfer between feature maps at different network levels, this improvement alleviates the loss of target information caused by changes in feature map resolution. The improved model can integrate more detailed information from multiple layers, enabling the network to better focus on the area containing small targets.

2.2.2 BiFPN_F strategy

In order to better integrate the features extracted by the backbone network, this paper retains the FPN + PAN structure in the Neck network^[24]. However, the

introduction of the P2 detection head leads to an increase in the number of model parameters. Additionally, the detailed features of the fire source targets are relatively few. The ‘‘Concat’’ feature maps merging method

increases the number of network channels, while the ‘‘Add’’ feature maps merging method keeps the number of channels unchanged, thereby reducing the overall model parameters.

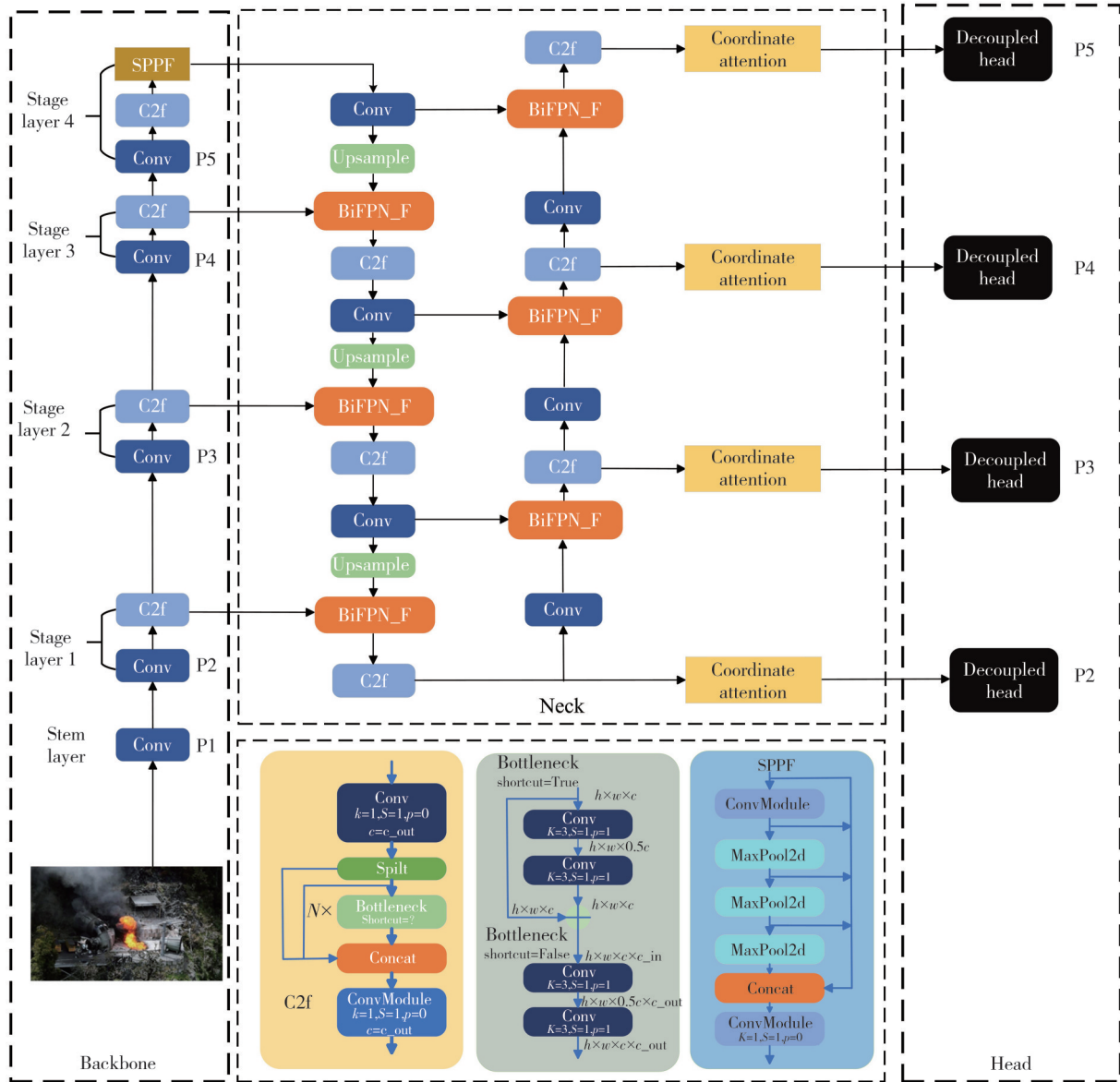


Fig. 2 Network structure of YOLO-fire model

To address the issue of decreased real-time performance caused by the increased parameter count in the neck network, this paper proposes a neck network structure based on BiFPN_F. By adopting a feature fusion strategy based on BiFPN_F for different feature layers, the model parameters are effectively controlled after incorporating the P2 detection head. The original neck and the BiFPN_F-based neck network structure are shown in Fig.3.

2.2.3 CA mechanism

Channel attention mechanisms (such as squeeze-and-excitation (SE) attention) have demonstrated significant effectiveness in enhancing model performance. However,

they primarily consider the importance of each channel while neglecting the crucial role of positional information in generating spatially selective attention maps. CA proposes a more effective approach to capture both positional information and channel relationships^[26]. This attention mechanism first decomposes channel attention into two one-dimensional feature encoding processes that aggregate features along the two spatial directions, respectively. One spatial direction captures long-range dependencies, while the other preserves precise positional information. Through feature encoding, a pair of direction-aware and position-sensitive attention maps is obtained. These attention maps are then merged into a single feature layer via transpose and

concatenation operations, followed by convolution (Conv2d), batch normalization, and non-linear activation to generate a new feature layer. Finally, a split operation is performed to recover the two directional feature layers. Using 1×1 convolution to adjust channel dimensions and the sigmoid function to obtain attention weights across

different dimensions, these are complementarily applied to the input feature map to enhance the representation of objects of interest. The structure of the CA mechanism is illustrated in Fig.4. C is the number of feature channels, W and H are the width and height of the feature map respectively.

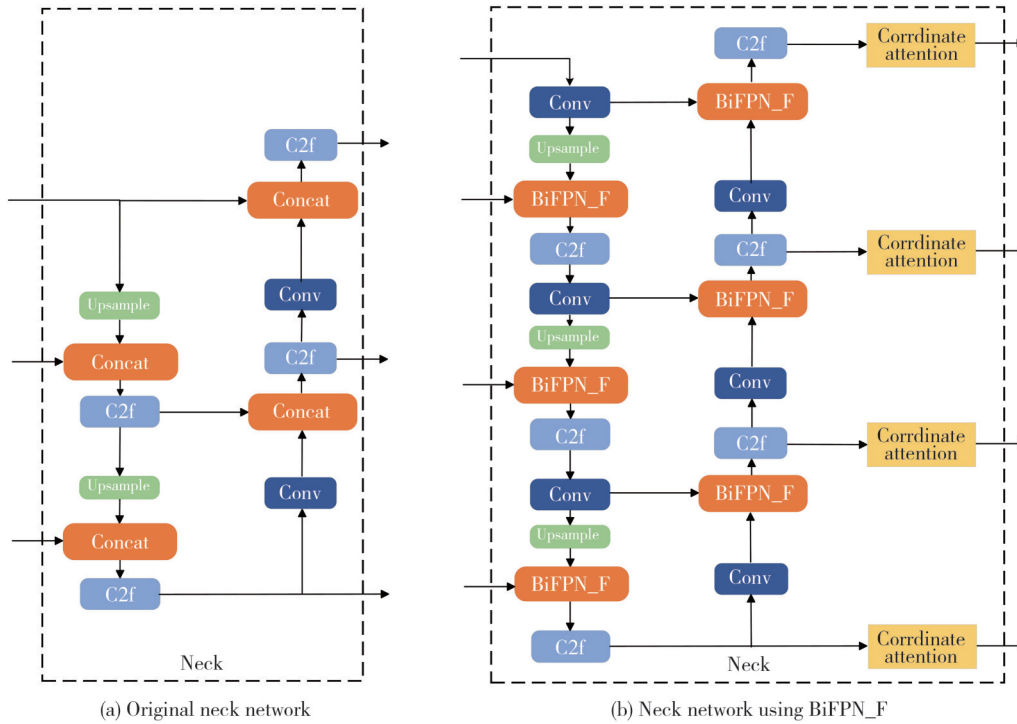


Fig. 3 Network structure of neck

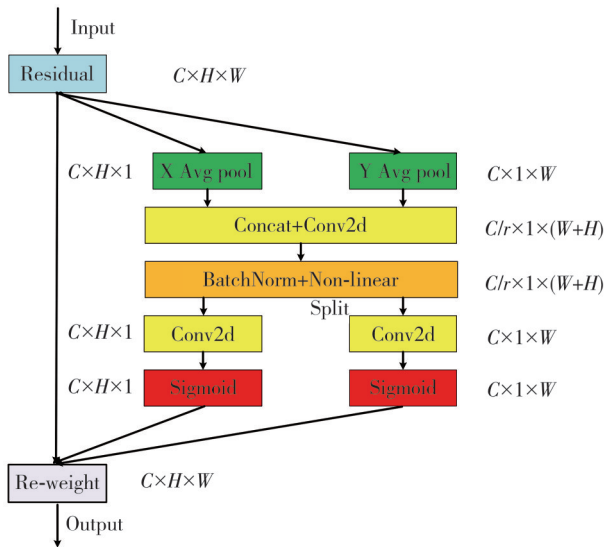


Fig. 4 Structure of CA mechanism

2.2.4 WIoU loss function

The loss function for object detection tasks generally consists of two components: classification loss and bounding box regression loss^[24]. When the commonly used regression loss function of intersection over union (IoU) equals 0, it only indicates that the predicted bounding box and the target box do not overlap, but fails

to reflect the distance between the two boxes. Furthermore, since IoU is non-differentiable when its value is 0, the IoU loss function cannot optimize scenarios where the two boxes do not intersect.

The improved loss functions include GIoU, CIoU, EIoU, and WIoU. Among these, the WIoU loss function introduces a dynamic weighting factor t during IoU calculation, enabling the model to focus more on hard samples. Since fire targets at different ignition stages exhibit diversity in scale and morphology, and fire source scenarios in confined spaces are complex with numerous occluded fire targets, this paper adopts the WIoU loss function. The WIoU loss function is

$$L_{\text{WIoU}} = t(1 - \text{IoU}), \quad t = \frac{e^{(1 - \text{IoU})}}{\gamma}, \quad (1)$$

where L_{WIoU} is loss function value, t is the dynamic weighting factor, γ is the hyperparameter, and IoU is the IoU value of the regression loss function.

3 Experiments

3.1 Experimental setup

The experiment was conducted on a Windows

operating system using PyTorch as the development framework. The experimental environment and configuration parameters are presented in Table 1.

Table 1 Experimental environment settings

Parameter	Configuration
CPU	AMD Ryzen 9 7940H w/Radeon 780M Graphics
GPU	NVIDIA GeForce RTX 4060 (12 GB)
CUDA version	Cuda 11.8
Python version	Python 3.10
Deep learning framework	PyTorch 2.3.0
Operating system	Windows 11

During model training, specific parameters and hyperparameters were used to ensure optimal performance. The detailed parameter settings are listed in Table 2. The selection of these hyperparameters is based on the results of multiple experiments and precedents from previous studies^[24], ensuring the model's excellent performance under various conditions.

Table 2 Model parameter settings

Parameter	Value
Training epochs	300
Batch size	16
Optimizer	SGD
Momentum	0.937
Weight decay	0.000 5
Learning rate	0.01

3.2 Dataset introduction

The diversity and quality of data samples are crucial for the accuracy of fire detection. Therefore, considering the impact of different environments and weather conditions on fire detection, this paper utilizes a self-built multi-scale fire dataset, comprising a total of 20 558 images. This dataset collects real-scene photos taken under various weather and lighting conditions, covering typical scenarios such as indoor cooking fires and outdoor waste burning activities at different times of the day, ensuring data richness to guarantee the broad applicability of the trained model across diverse environments. Additionally, to improve dataset quality, images with indistinct target features or excessive repetition were excluded. The dataset was further expanded with additional indoor and tunnel fire data samples, and all annotated using the LabelImg annotation tool. During annotation, complete flame regions were selected to ensure the integrity of the detected target features. Sample images are shown in Fig.5.

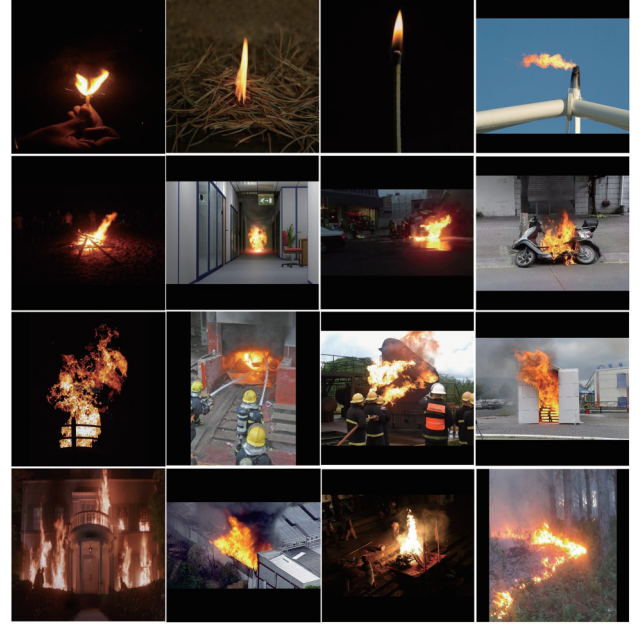


Fig. 5 Sample images from dataset

3.3 Evaluation metrics

To comprehensively evaluate the overall performance of the model, the experiment selected seven evaluation metrics: precision, recall, $F1$ -score, mean average precision (mAP), model size/MB, giga floating-point operations per second (GFLOPs), and parameters/M. They are calculated by

$$P = \frac{TP}{TP + FP} \cdot 100\%, \quad (2)$$

$$R = \frac{TP}{TP + FN} \cdot 100\%, \quad (3)$$

$$F1 = 2 \cdot \frac{P \cdot R}{P + R}, \quad (4)$$

$$mAP = \frac{1}{n} \sum_{k=1}^n AP_k \cdot 100\%, \quad (5)$$

where P is the precision of the model predictions, TP is the number of true positives, FP is the number of false positives, R is the recall rate of the model predictions, FN is the number of false negatives, $F1$ is the $F1$ -score value (the harmonic mean of P and R), mAP is the mean average precision across predicted categories, AP_k is the average precision for the k -th category, n is the number of categories.

3.4 Ablation studies

To validate the effectiveness of each designed and optimized module in YOLO-Fire, this paper systematically evaluates their impact on algorithm performance by incrementally adding modules. The specific configurations

for the ablation studies are as follows: ① Baseline model (YOLOv8n), ② YOLOv8n with only the P2 detection head introduced, ③ YOLOv8n incorporating only the BiFPN_F strategy, ④ YOLOv8n with the BiFPN_F

strategy and the CA mechanism, ⑤ YOLOv8n incorporating the P2 detection head, BiFPN_F strategy, CA mechanism, and replacing CIoU with the WIoU loss function. The experimental results are shown in Table 3.

Table 3 Experimental results of ablation studies

Model	Modules				Evaluation metrics						
	P2	BiFPN_F	CA	WIoU	P/%	R/%	mAP50/%	mAP50-95/%	Parameters/M	GFLOPs	Model size/MB
①					65.8	49.9	55.6	32.0	3.006	8.1	6.3
②	✓				65.9	49.7	55.6	32.2	2.921	12.2	6.3
③		✓			65.1	49.9	55.8	32.0	2.879	12.0	6.3
④		✓	✓		64.5	50.2	55.4	32.4	2.892	12.0	6.3
⑤	✓	✓	✓	✓	66.7	49.6	55.4	32.4	2.892	12.0	6.3

In fire detection, both *FP* and *FN* can lead to severe consequences. To more intuitively evaluate the performance of the YOLOv8n model and its optimized variants with different modules, as well as to assess the balance between precision and recall across different models, this paper substitutes the *P* and *R* values from Table 3 into Eq.3 to calculate the corresponding *F1*-scores and presents the results in the bar chart shown in Fig.6.

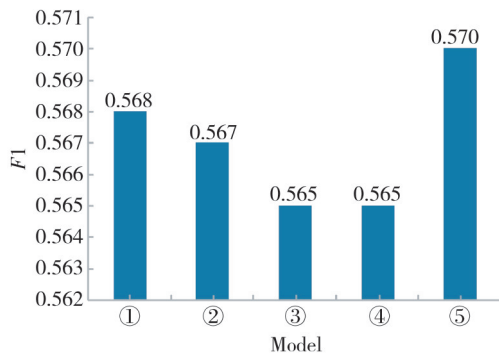


Fig. 6 F1-score values of different network models

As shown in Table 3 and Fig.6, compared to the baseline model ①, the model ② achieves a precision of 65.9%, an mAP50-95 of 32.2%, and a parameter count of 2.921 M, indicating that the model with the P2 detection head can more effectively detect small fire targets. After adopting the BiFPN_F strategy, the model's mAP50 shows significant improvement, while the parameter count decreases to 2.879 M, demonstrating that the BiFPN_F strategy effectively reduces model parameters while maintaining detection performance. Furthermore, when the neck network combines the BiFPN_F strategy with the CA module, the model's recall increases by 0.6% and the mAP50-95 improves by 1.25%, indicating that the CA module can more comprehensively capture dependencies between features, thereby enhancing the model's detection performance. Finally, on the basis of model ④, introducing the P2 detection head and replacing CIoU with WIoU results in stable recall, mAP, parameters, GFLOPs, and model size, while significantly improving detection

precision. This demonstrates that the WIoU loss function mitigates the neglect of special samples during training, and the introduction of the P2 detection head further enhances the model's ability to detect small targets.

Compared to the baseline model ①, the model ⑤ (YOLO-Fire) achieves improvements of 1.37% in precision, 1.25% in mAP50-95, and 0.35% in *F1*-score, while reducing parameters by 3.79%, indicating a comprehensive enhancement in overall performance. In conclusion, for multi-scale fire detection, the modules designed and introduced in this study are of significant importance, not only excelling in balancing precision and recall but also laying the foundation for deploying the model on edge devices.

3.5 Comparison studies

To validate that YOLO-fire's fire detection capability surpasses that of mainstream object detection models, this paper selects YOLOv5n^[27], YOLOv8n^[28], YOLOv10n^[29], RetinaNet^[30], EfficientDet^[31], CenterNet^[32], and RT-DETR^[33] algorithms for comparative analysis. The experimental results are presented in Table 4.

As shown in Table 4, the YOLO-fire model demonstrates optimal overall performance. Compared to the earlier YOLO version YOLOv5n, YOLO-fire not only significantly outperforms in precision, recall, mAP50, mAP50-95, and *F1*-score, but also maintains comparable model complexity, parameters, and model size. Relative to the baseline model, YOLO-fire achieves improvements of 1.37% in precision, 1.25% in mAP50-95, and 0.35% in *F1*-score, while reducing parameters by 3.79%. When compared to YOLOv10n, YOLO-fire shows enhancements of 1.99% in precision, 2.86% in mAP50-95, and 0.53% in *F1*-score.

To further validate the superiority of YOLO-fire, this study also conducted comparisons with other mainstream object detection models. Compared to the advanced lightweight algorithm EfficientDet, YOLO-

fire not only achieves comprehensive outperformance in recall, mAP50, mAP50-95, and *F1*-score, but also reduces parameters and model size by 25.35% and 58%, respectively. When compared to the Anchor-free detection algorithm CenterNet, YOLO-fire improves mAP50-95 and *F1*-score by 0.62% and 81.53%,

respectively. Relative to the single-stage algorithm RetinaNet, YOLO-Fire demonstrates superior performance in multi-scale fire detection, algorithmic complexity, and lightweight design. The model also exhibits significant advantages across all parameters when compared with the recently proposed RT-DETR.

Table 4 Comparison of mainstream models performance

Model	Evaluation metric							
	<i>P</i> /%	<i>R</i> /%	<i>mAP</i> 50/%	<i>mAP</i> 50-95/%	<i>F1</i>	Parameters/M	GFLOPs	Model size/MB
YOLOv5n	64.5	49.2	54.8	31.4	0.558	2.503	7.1	5.3
YOLOv8n	65.8	49.9	55.6	32.0	0.568	3.006	8.1	6.3
YOLOv10n	65.4	50.0	55.7	31.5	0.567	2.695	8.2	5.8
EfficientDet	76.5	40.6	51.6	30.2	0.530	3.874	5.2	15.0
CenterNet	93.4	18.9	56.0	32.2	0.314	32.665	70.2	124.0
RetinaNet	76.5	40.6	51.6	30.2	0.530	37.969	170.1	138.0
RT-DETR	55.8	43.8	51.2	29.9	0.491	31.988	103.4	66.2
Ours	66.7	49.6	55.4	32.4	0.570	2.892	12.0	6.3

To more intuitively demonstrate the superior performance of YOLO-fire, we normalize each evaluation metric and generate corresponding radar charts for each algorithm, as shown in Fig.7.

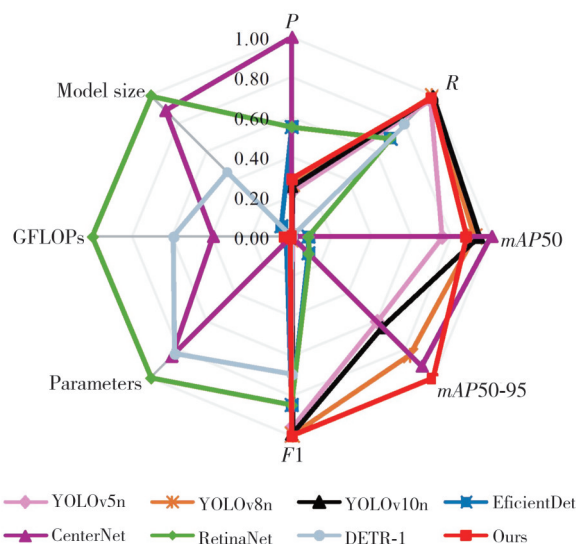


Fig. 7 Radar chart of normalized performance metrics

As clearly illustrated in Fig.7, compared to YOLO series models such as YOLOv5n, YOLOv8n, and YOLOv10n, the YOLO-fire maintains an overall leading advantage in performance metrics. When compared to existing mainstream object detection models, the YOLO-fire demonstrates superior performance in terms of model size, parameters, and algorithmic complexity. These results fully demonstrate our model's exceptional balance between performance and efficiency, thereby providing technical support for intelligent fire prevention and control across diverse scenarios.

For fire detection models deployed at the edge, we not only require high accuracy and recall (low rate of

missed detections) but also a lightweight model. To achieve this, the paper adopts the *F1*-score to comprehensively evaluate the model's detection performance. As shown in Table 4 and Fig.7, compared to the baseline model and mainstream models, the proposed model achieves the highest *F1* and *mAP* values, with more significant improvements.

3.6 Visual presentation

After 300 epochs of training on the self-built multi-scale fire, the performance metrics of the training set and validation set for the YOLO-fire model are shown in Fig.8.

The first three columns in Fig.8 depict the variation of the box_loss, cls_loss, and dfl_loss functions over time in the training and validation sets, respectively. The loss curves for both the training and validation sets show an initial rapid decline followed by convergence. In the early stages, the sharp decrease in the three loss functions within the training set reflects the model's strong adaptability to data characteristics during the learning process. As the number of iterations increases, the loss functions in the validation set quickly stabilize, indicating improved generalization capability of the model on unseen data. Furthermore, the loss curves exhibit a smooth and stable trend, demonstrating that overfitting is effectively controlled.

The last two columns in Fig.8 illustrate the performance curves of the evaluation metrics. As the training epochs increase, the evaluation metrics of the model rise rapidly and stabilize without significant fluctuations, indicating good learning stability and convergence of the model, with no overfitting observed during the training process. In

conclusion, the YOLO-fire demonstrates strong fitting capability and stability.

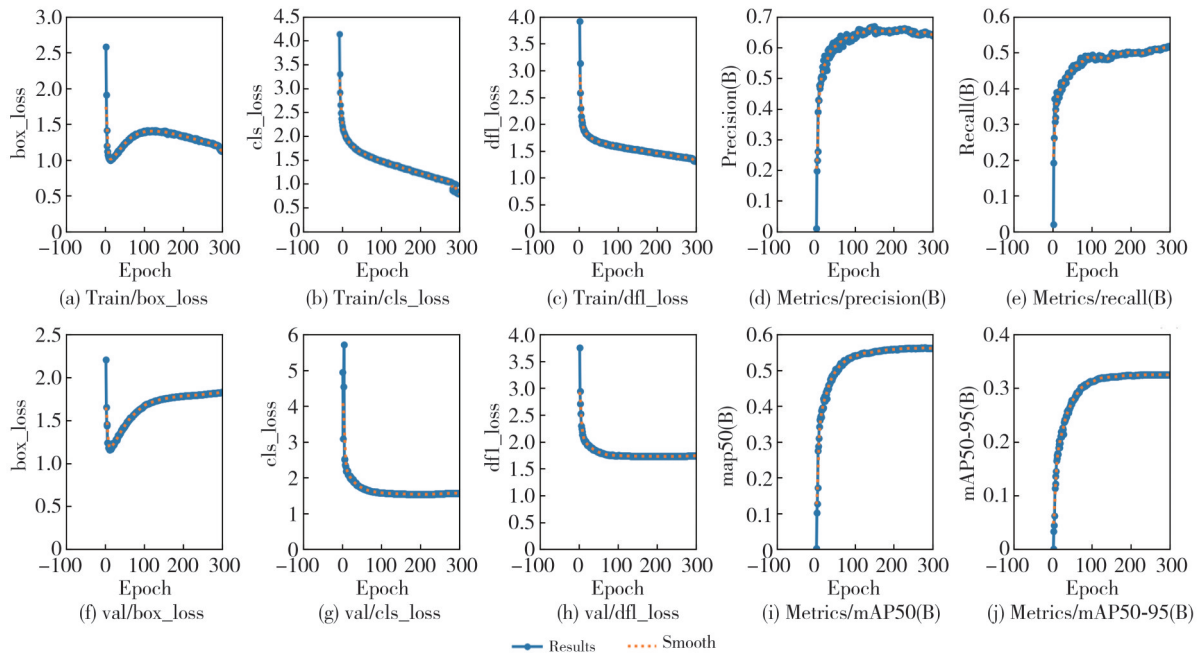


Fig. 8 Model training performance metrics curves

To demonstrate the powerful performance of YOLO-fire, we present the multi-scale fire target detection results corresponding to Fig.5, as shown in Fig.9. Each detection result is represented by a rectangular bounding box, accompanied by the target's category label and confidence score for clear identification, clearly showcasing the model's detection performance across varying scales.

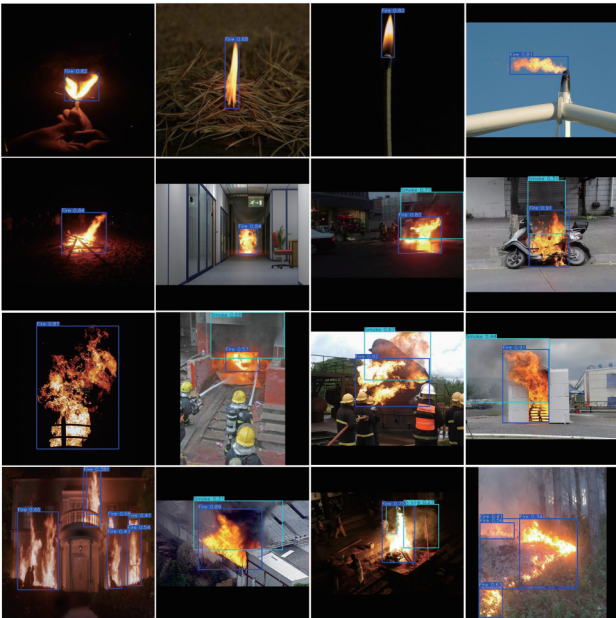


Fig. 9 Sample results by the model's detection

As observed in Fig.9 and validation results, although YOLO-fire exhibits occasional false positives and missed detections in certain scenarios, it maintains stable detection capabilities for fire targets across

multiple scales in diverse environments, highlighting its exceptional performance advantages and generalization ability. The model demonstrates particularly outstanding performance in detecting small-target fire incidents. This remarkable capability not only validates the model's reliability in practical applications but also establishes a foundation for its deployment on edge devices.

4 Conclusions

To address the challenges in fire monitoring technology, such as the highly variable morphology of fire sources and the significant differences in visual characteristics across different fire stages that complicate fire detection, this paper proposes a multi-scale fire target detection algorithm using YOLO-fire. By introducing a P2 detection head, the model's capability to detect early-stage fires is enhanced. Through the adoption of the BiFPN_F strategy for integrating features from adjacent layers, the cross-scale representation of fire characteristics is strengthened while reducing the number of model parameters. Additionally, by optimizing the network structure and introducing the CA mechanism, the model's ability to detect multi-scale fire targets and its scene generalization ability are further improved.

Ablation studies and comparative experiments demonstrate that YOLO-fire effectively balances detection performance, model lightweight design, and algorithmic complexity in existing target detection models. For multi-

scale fire target detection, compared to the baseline model, YOLO-fire achieves improvements of 1.37% in precision, 1.25% in mAP50-95, and 0.35% in *F1*-score, while reducing parameters by 3.79%. Furthermore, compared to current mainstream target detection models, YOLO-fire also achieves optimal detection performance.

The current research focuses solely on multi-scale fire detection. However, limitations of the self-built multi-scale fire dataset and the limited categories of ignition features prevent it from fully representing the complexity and diversity of different environments, thereby restricting the model's generalization capability across various scenarios. Future efforts should prioritize the collection of more diverse and complex fire data to validate and enhance the overall performance and scenario applicability of YOLO-fire.

Acknowledgement

This work was supported by Natural Science Foundation of Inner Mongolia Autonomous Region, China (No.2023QN05023), the Key R&D and Achievement Transformation Programs of Inner Mongolia Autonomous Region, China (Nos.2025KYPT0051, 2025KYPT0050).

Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

References

- [1] VASCONCELOS R N, FRANCA ROCHA W J S, COSTA D P, et al. Fire detection with deep learning: a comprehensive review. *Land*, 2024, 13(10): 1696.
- [2] ZHENG H T, WANG G Y, XIAO D, et al. FTA-DETR: an efficient and precise fire detection framework based on an end-to-end architecture applicable to embedded platforms. *Expert Systems with Applications*, 2024, 248: 123394.
- [3] ZHAO H Y, JIN J, LIU Y, et al. FSDF: a high-performance fire detection framework. *Expert Systems with Applications*, 2024, 238: 121665.
- [4] SUN J P, SUN Y Y, FAN W Q. Mine exogenous fire identification method based on visible light and infrared image. *Industry and Mine Automation*, 2019, 45(5): 1-5.
- [5] FAN W Q, LI X Y, LIU Y, et al. Mine external fire monitoring method using the fusion of visible visual features. *Journal of Mining Science and Technology*, 2023, 8(4): 529-537.
- [6] LI X Y, FAN W Q, LIU Y, et al. Mine exogenous fire monitoring method using the fusion of infrared visual features. *Journal of Mining Science and Technology*, 2025, 10(1): 116-124.
- [7] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 18-24, 2014, Columbus, Ohio, USA. New York: IEEE, 2014: 580-587.
- [8] GIRSHICK R. Fast R-CNN//2015 IEEE International Conference on Computer Vision, December 11-18, 2015, Santiago, Chile. New York: IEEE, 2015: 1440-1448.
- [9] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [10] YU L C, LIU J Q. Fire image recognition algorithm based on improved mask R-CNN. *Computer Engineering and Applications*, 2020, 56(21): 194-198.
- [11] ZHANG S L, ZHANG Y N, TIAN C, et al. Study on flame image recognition of chemical industrial park fires based on convolutional neural network. *China Safety Science Journal*, 2024, 34(1): 179-186.
- [12] BARMPOUTIS P, DIMITROPOULOS K, KAZA K, et al. Fire detection from images using faster R-CNN and multidimensional texture analysis//2019 IEEE International Conference on Acoustics, Speech and Signal Processing, May 12-17, 2019. Brighton, KingdomUnited. New York: IEEE, 2019: 8301-8305.
- [13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 779-788.
- [14] REDMON J, FARHADI A. YOLO9000: better, faster, stronger//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 6517-6525.
- [15] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [16] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021. (2021) [2024-10-08]. arXiv:2107.08430. <https://doi.org/10.1109/TPAMI.2022.3189759>
- [17] QIAN H M, SHI F, CHEN W, et al. A fire monitoring and alarm system based on channel-wise pruned YOLOv3. *Multimedia Tools and Applications*, 2022, 81(2): 1833-1851.
- [18] ZHAO L, ZHI L Q, ZHAO C, et al. Fire-YOLO: a small target object detection method for fire inspection. *Sustainability*, 2022, 14(9): 4930.
- [19] CHENG S H, MA J Y, ZHANG S J, et al. Smoke detection algorithm combined with improved Gaussian mixture and YOLOv2. *Acta Metrologica Sinica*, 2019, 40(5): 798-803.
- [20] FAN R X, PEI M T. Lightweight forest fire detection based on deep learning//2021 IEEE 31st International Workshop on Machine Learning for Signal Processing,

- October 25-28, 2021, Gold Coast, Australia. New York: IEEE, 2021: 1-6.
- [21] AKHMEDOV F, NASIMOV R, ABDUSALOMOV A. Dehazing algorithm integration with YOLO-v10 for ship fire detection. *Fire*, 2024, 7(9): 332.
- [22] HE N L, ZHANG J S, LIN W S. Research on forest fire image recognition based on deep learning multi-object detection technology. *Journal of Nanjing Forestry University (Natural Sciences Edition)*, 2024, 48(3): 207-218.
- [23] HE Y, HU J J, ZENG M, et al. DCGC-YOLO: the efficient dual-channel bottleneck structure YOLO detection algorithm for fire detection. *IEEE Access*, 2024, 12: 65254-65265.
- [24] LIU D, ZHAO X, FAN W Q. A small object detection algorithm for mine environment. *Engineering Applications of Artificial Intelligence*, 2025, 153: 110936.
- [25] CHU Y, WEI Y L, LIANG H G, et al. Lightweight YOLOv8 algorithm for weed target detection. *Journal of North University of China (Natural Science Edition)*, 2025, 46(4): 489-498.
- [26] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 20-25, 2021. Nashville, TN, USA. New York: IEEE, 2021: 13708-13717.
- [27] JOCHER G, STOKEN A, CHAURASIA A, et al. ultralytics/yolov5: v6.0—YOLOv5n “Nano” models, Roboflow integration, TensorFlow export, OpenCV DNN support. (2021) [2024-10-08]. <https://doi.org/10.5281/zenodo.5563715>
- [28] JOCHER G, CHAURASIA A, QIU J. YOLO by Ultralytics. (2023, Version 8.0.0) [2024-10-08]. <https://github.com/ultralytics/ultralytics>
- [29] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection. 2024: arXiv: 2405.14458. <https://arxiv.org/abs/2405.14458>
- [30] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2999-3007.
- [31] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 14-19, 2020. Seattle, WA, USA. New York: IEEE, 2020: 10781-10790.
- [32] DUAN K W, BAI S, XIE L X, et al. CenterNet: keypoint triplets for object detection//2019 IEEE/CVF International Conference on Computer Vision, October 27-November 2, 2019. Seoul, Korea. New York: IEEE, 2019: 6568-6577.
- [33] ZHAO Y A, LÜ W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 17-21, 2024, Seattle, WA, USA. New York: IEEE, 2024: 16965-16974.

基于 YOLO-fire 的多尺度火灾目标检测算法

范伟强^{1,2*}, 丁嘉彦¹, 彭斌¹, 刘冬¹, 高硕亨¹, 贾昌卓¹

1. 内蒙古大学电子信息工程学院, 内蒙古呼和浩特 010021;

2. 内蒙古自治区智慧通信感知与信号处理重点实验室, 内蒙古呼和浩特 010021

摘要: 火灾的精准和实时检测对工业生产和日常生活而言至关重要。然而, 火灾的形态不固定, 并且火灾不同阶段的视觉特征差异大, 使得火灾的精准防控面临着巨大挑战。为解决这一难题, 结合 YOLOv8 模型, 通过引入了新的层结构和注意力机制, 并替换新的特征融合模块和损失函数, 本文提出了基于 YOLO-fire 的多尺度火灾目标检测算法。通过引入小目标检测 P2 层, 提升模型对早期火灾的检测能力。引入坐标注意力机制并融入不同尺度目标检测的层结构中, 提升目标位置信息和通道关系的捕获能力, 使其更聚焦于目标区域。优化 Neck 网络结构, 对不同的特征层采用基于 BiFPN_F 的融合策略, 增强火灾特征的跨尺度表达能力, 并控制所设计模型的参数量。采用 WIoU 损失函数优化回归过程, 提升复杂场景下火源定位精度, 增强模型的鲁棒性并提高检测精度。在火灾数据集上的实验结果表明, YOLO-fire 能够有效检测不同场景下的多尺度火灾目标。相较于基准模型 YOLOv8n, YOLO-fire 的准确率、mAP50-95、F1 分数分别提升了 1.37%、1.25% 和 0.35%, 参数量下降了 3.79%。相较于当前主流目标检测算法, YOLO-fire 也实现了最优的检测性能, 并减少了网络参数和计算复杂度。该研究可为相关领域的火灾安全防护提供有效的技术支持。

关键词: 火灾检测; 早期火灾; 特征融合; 注意力机制; 损失函数; 网络结构; YOLOv8

引用格式: FAN Weiqiang, DING Jiayan, PENG Bin, et al. Multi-scale fire target detection algorithm using YOLO-fire. *Journal of Measurement Science and Instrumentation*, 2025, 16(4): 625-636. DOI: 10.62756/jmsi.1674-8042.2025060