Improved Strategy of Lightweight Fire Early Detection Algorithm Based on GBM-YOLO

Shuogui Zeng[1,a], Ye Chen[1,b,*], Jun Li[1,c], Tao Liu[1,d], Fuwei He[2,e], Zhenxi Liu[3,f]

[1]School of Mechanical Engineering, Sichuan University of Science & Engineering, No. 1, Baita Road, Sanjiang New Area, Yibin, Sichuan, China
[2]Luzhou Fire and Rescue department, Luzhou, Sichuan, China
[3]Yibin Fire and Rescue department, Yibin, Sichuan, China

[a]1187033047@qq.com, [b*]cy32428yr@suse.edu.cn,
[c]58501323@qq.com, [d]2242991685@qq.com,
[e]1377097764@qq.com, [f]330716347@qq.com

Abstract. Fire causes serious damage to buildings and infrastructure, threatening social security. Early fire feature detection can effectively reduce the harm caused by fire. However, traditional fire detection methods have problems such as difficult early detection, high false alarm rate, high missed detection rate and high maintenance cost. In order to solve the above problems, this paper proposes a lightweight fire detection model GBM-YOLO based on YOLOv8n. By designing a lightweight feature extraction network G-HGNetV2 to replace the backbone network of YOLOv8, the model complexity is reduced and the feature extraction ability is improved. A lightweight cross-scale adaptive weighted fusion module is designed in the neck network, and the DySample dynamic upsampling module is introduced to effectively fuse shallow and deep features to retain more feature details. The lightweight C2f-MPF module is designed to improve the feature extraction ability of the model for small fire targets and multi-scale targets. The results show that the average accuracy (mAP@50) of the improved model on the self-built fire fireworks data set reaches 86.1%, and the number of parameters (Params), computational complexity (GFLOPs) and model Size (Size) are reduced by 53.3%, 27.1% and 49.1%, respectively. The effect of the improved model in fireworks detection is improved. And the model is more lightweight and easier to deploy to edge devices to provide technical support for early warning of fire.

Keywords: fire detection; YOLOv8; Lightweight improvement; Small object detection

1 Introduction

Fire can cause damage to buildings and infrastructure, seriously affect economic development [1], and also pose a great threat to today's society. The characteristics of fire

K. Yuen et al. (eds.), *Proceedings of the 2025 International Conference on Engineering Management and Safety Engineering (EMSE 2025)*, Advances in Engineering Research 267,
https://doi.org/10.2991/978-94-6463-780-9_40

and smoke are random, and the early flame information is not obvious and difficult to be captured. Therefore, it is essential to carry out fire monitoring and real-time detection of early fires [2].

At present, fire identification mainly includes sensor detection technology and fireworks detection technology based on image recognition. Sensor detection technology is mainly based on the sensor capturing the chemical characteristics of flame, temperature and gas, which will generate an alarm signal when it reaches a certain threshold [3]. However, due to the limited sensitivity of traditional sensors, early detection of fire cannot be realized, and there are limitations in space and time. The interference caused by complex environment is prone to problems such as false alarms, inability to obtain real-time on-site images, and the need for regular maintenance and replacement [4]. Compared with sensor detection, image detection has the advantages of fast response time, wide detection range and low maintenance cost, so it has been widely concerned by researchers and fire protection enterprises.

Although existing fire image detection methods have achieved good accuracy, the detection speed is slow and relies on manual selection of flame and smoke characteristics [5]. In addition, due to the complexity of fireworks and interference factors in practical applications, pyrotechnic objects such as gas and light sources are difficult to distinguish from the flames and smoke in the actual scene.

In recent years, the rapid development of deep learning adopts more complex algorithms for automatic feature extraction and automatic detection to achieve faster learning, faster detection speed and higher accuracy, and is widely used in the field of fire detection. Although the improved model improves the detection accuracy to a certain extent, it greatly increases the amount of computation and is difficult to deploy to the equipment with limited resources. At present, the fire detection model does a good job in the detection of the fierce burning stage of the fire, but the detection performance of the weak flame and thin smoke in the early smoldering stage of the fire is not reliable, and it is difficult to detect the early warning signals of the fire, and it is more challenging to find these signals in the complex environment [6].

Therefore, how to detect the pyrotechnic information faster without bringing a huge computational burden to the network model, realize the lightweight design of the model, and deploy the target detection network to the edge device is a major problem facing the fire detection at present. Based on YOLOv8n model, this paper proposes a lightweight fire detection model - GBM-YOLO, and analyzes the detection effect of flame and smoke under different conditions, which provides guarantee and technical support for early fire warning, personal safety and property safety.

2 GBM-YOLO Fire Detection Algorithm

2.1 Lightweight Improvements Based on HGNetV2

In this paper, based on YOLOv8n deep learning algorithm, the benchmark model adopts the parameter sharing backbone network design, and uses a large number of 3×3 convolutional modules for feature extraction operation, which improves the overall generalization of the model, but it will lead to increased computing load, poor

computing power of edge devices, and delay of fireworks information detection. In order to solve the above problems, In this paper, GhostConv[7] is used in the backbone network of fireworks detection. This module performs feature mapping with cheap operations to generate additional feature maps, thereby reducing the complexity of the model.

In order to enhance the redundancy features and enrich the deep-level features of the backbone network, the backbone network HGNetV2[8] is combined with the GhostConv module, and the optimized backbone network is G-HGNetv2, as shown in Figure 1.

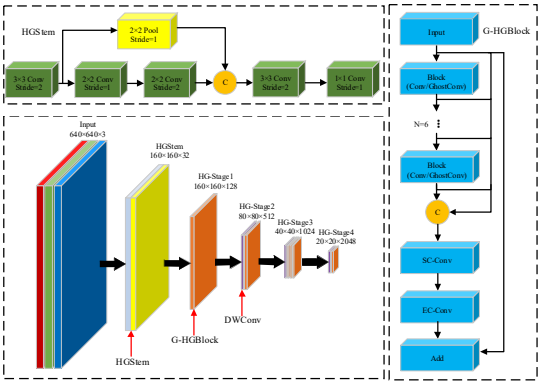


Fig. 1. Network structure of G-HGNetv2.

2.2 Reconstruction Design of Neck Network

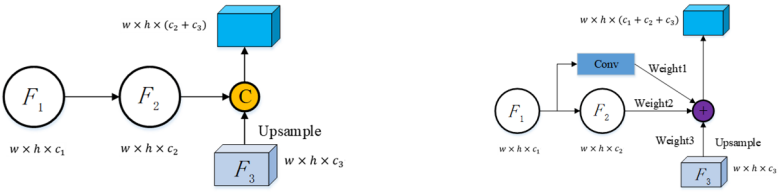


Fig. 2. a. Original feature fusion; b. Weighted feature fusion

In order to solve the problems of small size, large variation and low resolution of fireworks in fireworks detection tasks, this paper strengthened the ability of feature expression between small target information and different sizes based on the idea of multi-branch assisted FPN (MAFPN) [9]. BiFPN[10] can learn the weight fusion of information in deep feature map and shallow feature map to make up for the lack of detail information. Combined with MAFPN, BIFPN [10] is designed to alleviate the semantic interference of cross-scale feature fusion, realize the adaptive harmonization of features, enhance the module's ability to extract multi-scale features, and retain more

feature details. Finally, the stability of training is guaranteed by normalization processing. The weighted feature fusion method is shown in Figure 2 (b).

In Figure 2, C represents the Concat operation, + represents the add operation, and Weight1, Weight2 and Weight3 represent the learnable weight parameters of the corresponding feature graph. When layer i is fused, the feature connection formula of the middle layer and the output layer of the same layer is shown in (1-2)

$$P_i^{mid} = \mathrm{Conv}\left(\frac{\omega_1 \cdot P_i^{in} + \omega_2 \cdot \mathrm{Resize}\left(P_{i+1}^{mid}\right)}{\omega_1 + \omega_2 + \epsilon}\right) \tag{1}$$

$$P_i^{out} = \mathrm{Conv}\left(\frac{\omega_1' \cdot P_i^{in} + \omega_2' \cdot P_i^{mid} + \omega_3' \cdot \mathrm{Resize}\left(P_{i-1}^{out}\right)}{\omega_1' + \omega_2' + \omega_3' + \epsilon}\right) \tag{2}$$

In the formula, ω_i represents the learnable weight parameter, and the initial value of the weight coefficient is set to 1; The learning rate ϵ is set to 0.0001 to avoid instability. P_i^{mid} is the characteristic information of the middle layer. P_i^{out} is the output characteristic information; Resize is an operation that matches the resolution.

However, in the original up-sampling module, the sampling operator has a small field of perception, which leads to the loss of early fireworks information in the process of feature fusion, and the low quality of the image after up-sampling. In this paper, the dynamic up-sampling module DySample[11] is used to combine with the cross-scale weighted feature fusion network to avoid the boundary ambiguity problem that may arise in traditional interpolation methods and improve the model's ability to capture detailed features. Finally, the channel alignment operation is used to reduce the redundant channel information existing in the cross-scale connection structure. Based on the original model, the BMDFPN feature fusion network is designed, as shown in Figure 3.

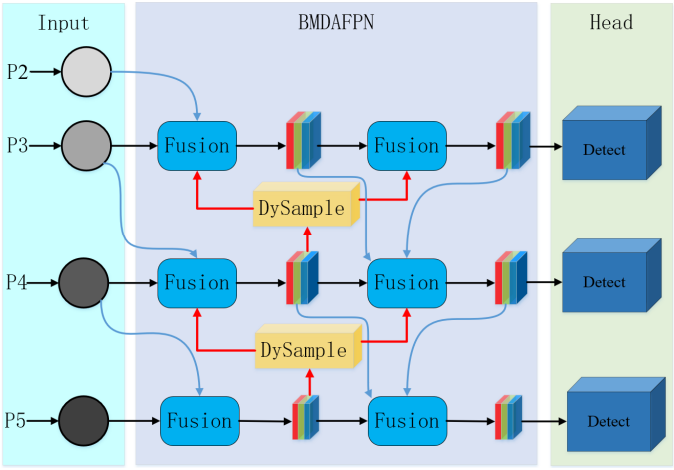


Fig. 3. BMDFPN.

2.3 C2f-MPF

In order to enhance the deep network's ability to express the information features of small fireworks and solve the problem of large changes in the size of fireworks, this paper proposes the C2f-MPF multi-scale local feature aggregation module, and the structure of the C2f-MPF module is shown in Figure 4. This module extends the receptive field without introducing additional parameters to better capture contextual information and preserve detailed features. The original bottleneck uses two 3×3 conventional convolutions in series for feature extraction, which will lead to multi-scale information loss and channel redundancy. MPF performs group convolution on half of the input feature map, and uses 3×3 and 5×5 convolutional layers of different sizes respectively, retains the other half of the feature map, and then uses a 7×7 convolutional layer to expand its receptive field to better capture the context information, and then performs Concat operation on the other half of the original features at each stage to avoid losing a certain scale after operation. The redundant information in the feature map is filtered, and then a 1×1 convolutional layer is used to interact with the feature information at different scales. Finally, the input features are residual concatenated with the processed features. The C2f-MPF module captures richer context information by gradually extracting multi-scale features and combining partial convolution, which is suitable for small object detection in complex scenes and significantly improves the adaptability of the model to multi-scale objects.

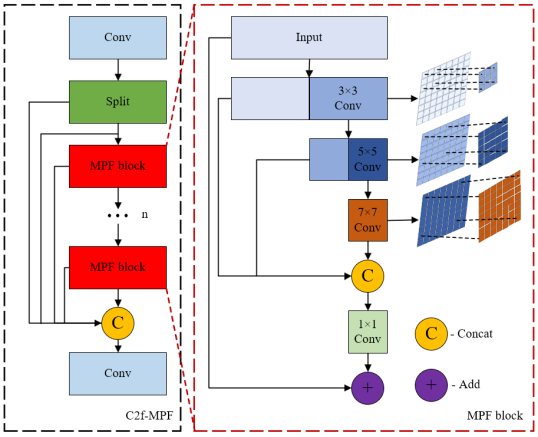


Fig. 4. C2f-MPF module.

2.4 Dataset

In order to improve the accuracy of deep learning, this paper integrates multiple public data sets [12-14] to establish a fire data set with a complex combustion environment and a large number of early pyrotechnic data, with a total of 16200 data.

The data set was divided into training set, verification set and test set according to the ratio of 8:1:1, and Labelimg tool was used to label all the fire and smoke information in

the data set diagram. The target size distribution diagram of the data set was shown in Figure 5, and the size of target objects was mainly concentrated in small targets, ensuring effective detection of the algorithm in the early stage of fire.

In order to evaluate the effectiveness of the model in fire detection, the average accuracy of the evaluation indexes commonly used in deep learning (mAP@0.5, mAP@0.5:0.95) was used in this experiment to evaluate the accuracy of the model, and the complexity of the model was evaluated by parameter number (Parms) and computational effort (GFLOPs).

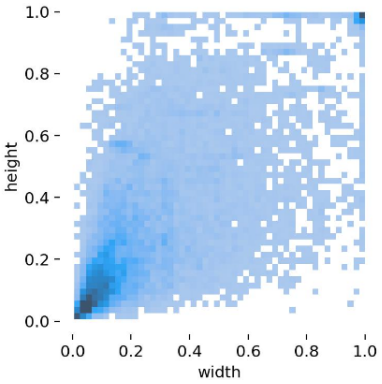


Fig. 5. Target size distribution map.

3 Comparative Analysis of Experimental Results

3.1 Performance Analysis of Backbone Network Reconstruction

Table 1 shows the performance comparison between this backbone network module and other feature extraction backbone network modules. As can be seen from the experimental results in Table 1, the reconstructed backbone network maintains the average accuracy, but the number of parameters and calculation amount decrease by 23.3% and 16%. Note The low-complexity feature mapping mode can effectively utilize the redundant information in the channel, retain rich feature expression capabilities, and reduce the computing load in lightweight deployment scenarios.

Table 1. Comparison experiments of backbone networks.

Base Model	*Backbone network*	$mAP_{0.5}$ *(%)*	$mAP_{0.5:0.95}$ *(%)*	*Params (M)*	*GFLOPs (G)*
YOLOv8n	BottleNeck	85.5	59.7	3	8.1
	EIEStem	84.6	59.7	3.0	8.3
	Starnet	84.7	58.4	2.2	6.5
	HGNetV2	85.4	59.3	2.4	6.9
	G- HGNetV2	85.5	59.7	2.3	6.8

Table 2 shows the results of various cervical network ablation experiments. As can be seen from Table 2, experiments show that MAFPN adopts weighted fusion and dynamic up-sampling modules to retain more detailed features on the basis of cross-scale connection, and finally compacts channel redundancy information, reducing the number of parameters and the amount of computation to 2.1M and 7.4G respectively.

Table 2. Comparison experiments of neck networks.

Base Model	*Neck network*	$mAP_{0.5}$ *(%)*	$mAP_{0.5:0.95}$ *(%)*	*Params (M)*	*GFLOPs (G)*
YOLOv8n	PAFPN	85.5	59.7	3	8.1
	MAFPN	85.4	59.6	2.9	8.7
	BIFPN	85.1	59.4	3.0	8.1
	AFPN	85.5	59.7	3.0	11.0
	MBDFPN	85.6	59.7	2.1	7.4

3.2 Comparison Experiment of Different C2f Structures

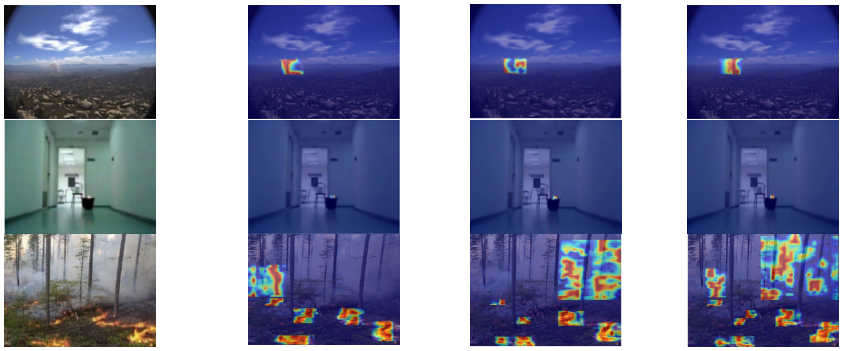
Lightweight C2f network modules C2F-faster [15], C2F-GCONV [16] and C2F-MOGA [17] are selected to verify the effectiveness of C2f module. Table 3 shows the comparative experiments of different C2f structures. It can be seen from the experimental results in Table 3 that all lightweight C2f modules lead to a decrease in the mean average accuracy, while C2F-MPF module can perceive multi-scale field of view while partially convolution, capture local field of view and global context information at the same time, and can handle more complex small target scenes. Compared with the C2f-Faster module which also adopts partial convolution, the feature extraction capability is more prominent, and more abundant gradient flow information can be obtained while ensuring lightweight.

Table 3. Comparison experiment with different C2f structures.

Base Model	*C2f module*	$mAP_{0.5}$ *(%)*	$mAP_{0.5:0.95}$ *(%)*	*Params (M)*	*GFLOPs (G)*
YOLOv8n	C2f	85.5	59.7	3	8.1
	C2f-Faster	84.6	57.6	2.3	6.3
	C2f-gConv	83.5	55	2.2	6.2
	C2f-Moga	83.3	54.7	2.7	7.6
	C2f-MPF	86.0	59.8	2.6	7.1

Fig. 6 shows the effect comparison of different improved module heatmaps. It can be seen from Fig. 6 (b) that the original model has a large area of missed detection of smoke detection, and the attention of small target fireworks is not focused. From Figure 6 (c), it can be seen that the C2f-Faster module has the problem of scattered attention when dealing with small target smoke, and there is also the problem of large-area missed detection of cross-scale target smoke. It can be seen from Fig. 6 (d) that in the scene with smoke like background interference, C2f-MPF module focuses its attention on the

smoke area, and its attention field and edge details are more complete compared with the other two modules. In the scene with small target flames, the attention to the flames is more obvious, and in the scene with cross-scale targets, the information is more complete. The experimental results show that the improved C2f-MPF module can capture small targets and cross-scale information of fireworks more comprehensively in complex environments, and improve the overall feature extraction ability and detection accuracy of the model.



(a). Original picture **(b).** YOLOv8n **(c).** C2f-Faster **(d).** C2f-MPF

Fig. 6. Comparison of thermal maps of different C2f modules.

3.3 Ablation Experiments of the Improved Model

In order to further verify the effect of the GMP-YOLO algorithm improvement strategy on the improvement of model detection ability, the detection performance of different modules proposed in this paper is ablated based on YOLOv8n. The experimental results are shown in Table 4, where "√" indicates the use of this module.

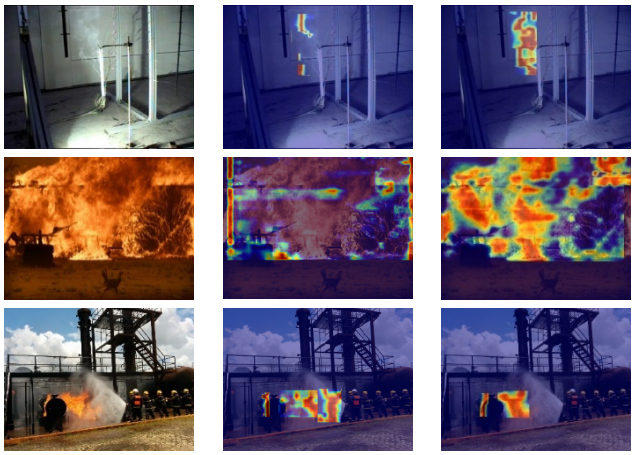
Table 4. Result of ablation experiments.

G-HGNetv2	*BMD-FPN*	*C2f-MPF*	*$mAP_{0.5}$ (%)*	*$mAP_{0.5:0.95}$ (%)*	*Params (M)*	*GFLOPs (G)*	*Size (MB)*
			85.5	59.7	3	8.1	5.96
√			85.5	59.7	2.3	6.8	4.65
	√		85.6	59.7	2.1	7.4	4.39
		√	86.0	59.8	2.6	7.1	5.21
√	√		85.6	59.7	1.4	6.2	3.08
√	√	√	**86.1**	**6.0**	**1.4**	**5.9**	**3.03**

According to the experimental results in Table 4, it can be seen that all the modules proposed in this paper can reduce the amount of model parameters and calculation. Compared with the baseline model, the model with all the improvements can reduce the amount of parameters, calculation and model size by 53.3%, 27.1% and 49.1%, respectively. mAP@0.5 and mAP@0.5:0.95 are increased by 0.6%and 0.3%respectively, indicating that the GBM-YOLO model can handle fireworks with complex backgrounds

and perform well in model lightweight, which makes it have better advantages in deploying edge devices.

Fig. 7 shows the comparison of visual heat maps. From Fig. 7 (b), it can be seen that the original model has no obvious perception of thin smoke in complex environments and the focus is scattered under cross-scale backgrounds, which can accurately reflect the target characteristics, and there is a false detection of smoke-like objects. From Figure 7 (c), it can be seen that the GBM-YOLO fireworks detection model can well reflect the characteristics of fireworks under complex backgrounds. In summary, the GBM-YOLO fireworks detection model in this paper can pay more attention to fireworks in complex backgrounds in an environment with limited computing resources.



(a). Original picture **(b).** YOLOv8n **(c).** GBM-YOLO

Fig. 7. Comparison of visual heat maps.

3.4 Comparison and Analysis of Different Algorithms

Table 5. Comparison of experimental results of different algorithms.

Id	*Network*	$mAP_{0.5}$ *(%)*	$mAP_{0.5:0.95}$ *(%)*	*Params (M)*	*FLOPs (G)*	*Size (Mb)*
1	yolov3-tiny	77.4	45.0	12.1	18.9	23.2
2	yolov5n	82.3	50.7	2.5	7.1	5.30
3	yolov6n	85.0	59.4	4.2	11.8	8.27
4	yolov8n	85.5	59.7	3.0	8.1	5.96
5	yolov9t	85.1	59.4	1.9	7.6	4.41
6	yolov10n	84.1	57.3	2.3	6.5	5.80
7	yolov11n	85.3	59.8	2.5	6.3	5.20
8	yolov12n	85.3	59.4	2.5	6.3	5.25
9	Hyper-YOLO	85.4	59.5	2.6	7.6	5.38
10	GBM-YOLO（ours）	**86.1**	**60.0**	**1.4**	**5.9**	**3.03**

In order to verify the effectiveness of the proposed algorithm, the experimental results of GBM-YOLO algorithm are compared with the current mainstream algorithms, and the results are shown in Table 5. The improved model proposed in this paper is superior to other algorithms in the accuracy of fire detection, and has the advantage of lightweight model structure, which makes it more suitable for fire detection on edge devices.

Figure 8 is a comparison chart of parameters and average accuracy of various models. It can be seen from Figure 8 that the abscis is the parameters and the ordinate is the average accuracy. The leftmost position of the improved GBM-YOLO model indicates that the model has the least parameters and simpler model structure, and the highest position indicates that the model has the highest accuracy of fireworks detection. It shows that the model applied in the actual fire detection can effectively protect people's life and property safety.

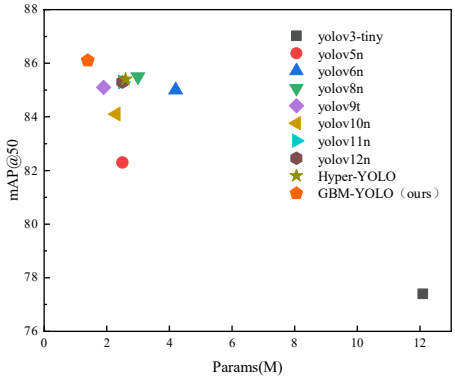


Fig. 8. Comparison of the number of parameters and average accuracy of various models.

4 Conclusions

(1) Through the analysis of experimental results, compared with the traditional algorithm, the G-HGNetV2 model proposed in this paper has the same accuracy as the traditional backbone network, and the number of parameters and calculation is reduced by 23.3% and 16%. The BMDFPN lightweight neck network model can retain more detailed features, and the number of parameters is reduced by 30%. The C2f-MPF module can effectively enhance the extraction ability of small targets and multi-scale features and reduce the computational complexity.

(2) Under the self-made data set, the GBM-YOLO algorithm mAP@50 reaches 86.1%, and the model parameters, calculation amount and model size are only 1.4M, 5.9G and 3.03MB, which are 53.3%, 27.1% and 49.1% lower than those of YOLOv8n, respectively. It shows the accuracy of the model in fireworks detection under complex background, and shows a good application prospect in the deployment of edge computing devices.

Acknowledgments

This work was financially supported by the Talent Introduction Project of Sichuan University of Science & Engineering (2019RC18), the Key Laboratory of Process Equipment and Control Engineering in Colleges and Universities of Sichuan Province (GK201910), and the 2024 Luzhou Science and Technology Program Projects (2024SYF183). Additionally, it was supported by the Scientific Research and Innovation Team Program of Sichuan University of Science & Engineering (SUSE652A004) and Yibin city science and technology plan project (2023SF007).

References

1. Xu, R.; Yu, P.; Abramson, M.J.; Johnston, F.H.; Samet, J.M.; Bell, M.L.; Haines, A.; Ebi, K.L.; Li, S.; Guo, Y. Wildfires, global climate change, and human health. N. Engl. J. Med. 2020, 383, 2173–2181.
2. Jiang X, Xu L, Fang X. DG-YOLO: A Novel Efficient Early Fire Detection Algorithm Under Complex Scenarios[J]. Fire Technology, 2024: 1-25.
3. Avazov K, Mukhiddinov M, Makhmudov F, et al. Fire detection method in smart city environments using a deep-learning-based approach[J]. Electronics, 2021, 11(1): 73.
4. He Y, Hu J, Zeng M, et al. DCGC-YOLO: The Efficient Dual-Channel Bottleneck Structure YOLO Detection Algorithm for Fire Detection[J]. IEEE Access, 2024.
5. Ye S, Bai Z, Chen H, et al. An effective algorithm to detect both smoke and flame using color and wavelet analysis[J]. Pattern Recognition and Image Analysis, 2017, 27: 131-138.
6. Yin H, Chen M, Fan W, et al. Efficient smoke detection based on YOLO v5s[J]. Mathematics, 2022, 10(19): 3493.
7. Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1580-1589.
8. Zhao Y, Lv W, Xu S, et al. Detrs beat yolos on real-time object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 16965-16974.
9. Yang Z, Guan Q, Zhao K, et al. Multi-branch Auxiliary Fusion YOLO with Re-parameterization Heterogeneous Convolutional for Accurate Object Detection[C]//Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Singapore: Springer Nature Singapore, 2024: 492-505.
10. Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10781-10790.
11. Liu W, Lu H, Fu H, et al. Learning to upsample by learning to sample[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 6027-6037.
12. Chino D Y T, Avalhais L P S, Rodrigues J F, et al. Bowfire: detection of fire in still images by integrating pixel color and texture analysis[C]//2015 28th SIBGRAPI conference on graphics, patterns and images. IEEE, 2015: 95-102.
13. Cetin, E.: Computer vision based fire detection dataset (May 2014), http://signal.ee.bilkent.edu.tr/VisiFire/
14. Mivia: Mivia fire detection dataset (May 2014), http://mivia.unisa.it/

15. Chen J, Kao S, He H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 12021-12031.
16. Song Y, Zhou Y, Qian H, et al. Rethinking performance gains in image dehazing networks[J]. arXiv preprint arXiv:2209.11448, 2022.
17. Li S, Wang Z, Liu Z, et al. Moganet: Multi-order gated aggregation network[C]//The Twelfth International Conference on Learning Representations. 2023.