Journal of Computers Vol. 36 No. 6, December 2025, pp. 21-33
https://doi.org/10.63367/199115992025123606002

Research on Forest Fire Detection Method Based on YOLOv12 Model

Chengmao Zhang[1*], Hongtu Zhang[2], and Yupeng Wei[2]

[1] College of Mechanical and Vehicle Engineering, Linyi University,
Linyi City 276000, Shandong Province, China
lydxzcm@126.com

[2] College of Automation and Electrical Engineering, Linyi University,
Linyi City 276000, Shandong Province, China
Zhanghongtu0221@126.com, 3310901287@qq.com

Received 21 March 2025; Revised 9 April 2025; Accepted 10 April 2025

Abstract. In the field of forest fire detection, accurately identifying smoke and flames has always posed a technical challenge. This study addresses this issue by proposing an improved model based on YOLOv12. The model adopts lightweight design principles and feature enhancement techniques, significantly improving detection accuracy and efficiency, and providing solid technical support for early warning of forest fires. After selecting YOLOv12 as the base architecture, the research team further integrated the AKConv variable kernel convolution module and the CBAM attention module, and incorporated the BiFPN weighted bidirectional feature pyramid network. The combination of these technologies aims to optimize the feature fusion process and significantly enhance the model's ability to detect small targets. The improved model achieved an average precision (mAP) of 72.8% on a self-built dataset, with a 7% increase in detection speed compared to the original model, and its robustness in complex scenarios has been significantly enhanced. This strategy not only improves detection accuracy through dynamic feature enhancement and efficient cross-scale fusion techniques, but also effectively reduces redundancy in the computation process, making it possible to build an "air-ground-space" integrated monitoring network, and promoting the shift towards intelligent early warning in forest fire prevention efforts.

Keywords: YOLOv12, forest fire detection, AKConv variable kernel convolution module, CBAM attention mechanism, BiFPN feature pyramid network

1 Introduction

Forest fires are a type of natural disaster that is sudden, highly destructive, and difficult to control. In recent years, the intensification of global climate change has led to their increased frequency, posing a severe threat to ecological security and human life and property. The study found that during the period from 1948 to 2014, the average number of fires per decade in the llmensky Reserve (southern Urals, Russia) increased by 1.9 times [1]; in 2019, the total area of burned land in the Beni Province of the Amazon region was approximately 1.18 million hectares, and the total area of burned land within or across the equatorial cities was about 3.82 million hectares [2]. These cases highlight the urgency of early fire detection. Traditional detection methods such as watchtowers and cameras have issues with small coverage areas and high false alarm rates, making them difficult to adapt to the complex and ever-changing forest environment. With the development of deep learning technology, fire detection methods based on object detection algorithms have gradually become a research hotspot. Among these, the YOLO series models have shown significant advantages in flame and smoke detection due to their real-time and high-precision characteristics [3]. However, smoke and flame targets in the early stages of forest fires are usually small in size and variable in shape. Existing models still lack sufficient detection accuracy and speed in complex scenarios, and there is an urgent need to improve performance through algorithm optimization.

Scholars both domestically and internationally have made significant progress in the field of fire detection based on deep learning. The applications of deep learning in fire detection are very rich. By comparing existing algorithms such as Faster-RCNN, R-FCN, SSD, and YOLO v3, it can be found that the fire detection algorithm based on YOLO v3 has a higher accuracy rate than other algorithms, and its detection speed is also good [4].

* Corresponding Author

Given the small size of fire points captured by drones, which often leads to mis-detection and missed detection during the detection process, a multi-scale feature fusion strategy is introduced to address the issue of weak small target detection capabilities in models. Based on YOLOv7, a lightweight small target detection model, FL-YOLOv7, is proposed [5]. Furthermore, in complex environments, issues such as missed and false detections in forest fire detection are particularly common. YOLOv8 has shown significant improvements compared to YOLOv7, and detection speed should also be enhanced [6]. Furthermore, when comparing YOLOv9, YOLOv10, and YOLOv11 for the detection of smoke and fire in forest environments, YOLOv11 demonstrates superior performance in terms of detection accuracy, recall, and mean Average Precision (mAP) [7]. Recently, new generation models such as YOLOv12 have further balanced accuracy and efficiency by simplifying the self-attention mechanism and optimizing feature fusion. However, existing methods still face challenges such as complex background interference, small object missed detection, and large model computational requirements, especially in real-time detection scenarios where both lightweight and high-precision demands need to be taken into account.

In this paper, an improved method for forest fire detection based on the YOLOv12 model is proposed. This method focuses on the accuracy and speed bottleneck issues present in current forest fire detection technologies and proposes a series of innovative solutions. Specifically, the main innovations are as follows:

(1) The introduction of the AKConv (Adaptive Kernel Convolution) module, which enhances the model's ability to capture multi-scale smoke and flame features by dynamically adjusting the shape and scale of the convolution kernels, while reducing computational redundancy.

(2) The integration of the CBAM (Convolutional Block Attention Module), which strengthens the extraction of key features and suppresses background interference through channel and spatial attention mechanisms.

(3) The adoption of the BiFPN (Bi-directional Feature Pyramid Network) structure, which achieves efficient cross-scale feature fusion through learnable feature weights, enhances the detection of small objects, and strengthens multi-scale information interaction through the repeated stacking of bi-directional paths.

Experimental results indicate that the improved YOLOv12 model, under the synergistic effect of the AKConv, CBAM, and BiFPN modules, achieves a mean Average Precision (mAP) of 78.8% at a threshold of 0.5, which represents a 2.9 percentage point increase compared to the baseline model. Additionally, the detection speed has been effectively enhanced by a 7% increase. In terms of error detection rates, the false positive rate has decreased by 14.2%, and the miss rate has been reduced by 9.8%. These enhancements significantly bolster the system's robustness in extreme scenarios, allowing for more stable and accurate execution of object detection tasks when faced with complex and variable environments.

2 YOLOv12 Model

The YOLO framework was proposed by Joseph Redmon, employing a grid-based approach to predict bounding boxes and class probabilities, revolutionizing real-time object detection. Its design enhances detection efficiency, making YOLO highly suitable for real-time applications. Since its inception, the YOLO architecture has undergone continuous upgrades to address various object detection challenges. Each update is dedicated to reducing latency and enhancing detection accuracy, which is crucial for real-time object detection technology [8]. The YOLO algorithm divides the image into grid cells, predicts bounding boxes and class probabilities, and performs object detection quickly and accurately. It utilizes convolutional layer features, extracting features through a CNN. To accommodate different object sizes, it introduces an anchor box mechanism and non-maximum suppression techniques to optimize detection results. The YOLO algorithm is continuously updated, incorporating new technologies to enhance detection speed and accuracy [9].

YOLOv12 algorithm is a popular object detection framework that offers five different versions: YOLOv12n, YOLOv12s, YOLOv12m, YOLOv12l, and YOLOv12x, to accommodate various application scenarios. Each version has its specific design to meet different needs ranging from lightweight to high performance. YOLOv12n, as one of the foundational models in the YOLOv12 series, maintains high accuracy while also providing fast detection speeds. Due to its unique features, YOLOv12n is particularly well-suited for detection in the field of fires, YOLOv12n Network Architecture Diagram (see Fig. 1). In addition to rapid detection and high accuracy, YOLOv12n also possesses excellent generalization capabilities, which means it can effectively detect new scenes and targets without extensive training. This feature makes YOLOv12n particularly important in emergency application scenarios such as fire detection, as it can quickly adapt to different fire environments and conditions, providing timely and reliable detection results. Furthermore, YOLOv12n employs advanced loss functions, further optimizing detection performance and ensuring stability and accuracy in various complex scenarios.

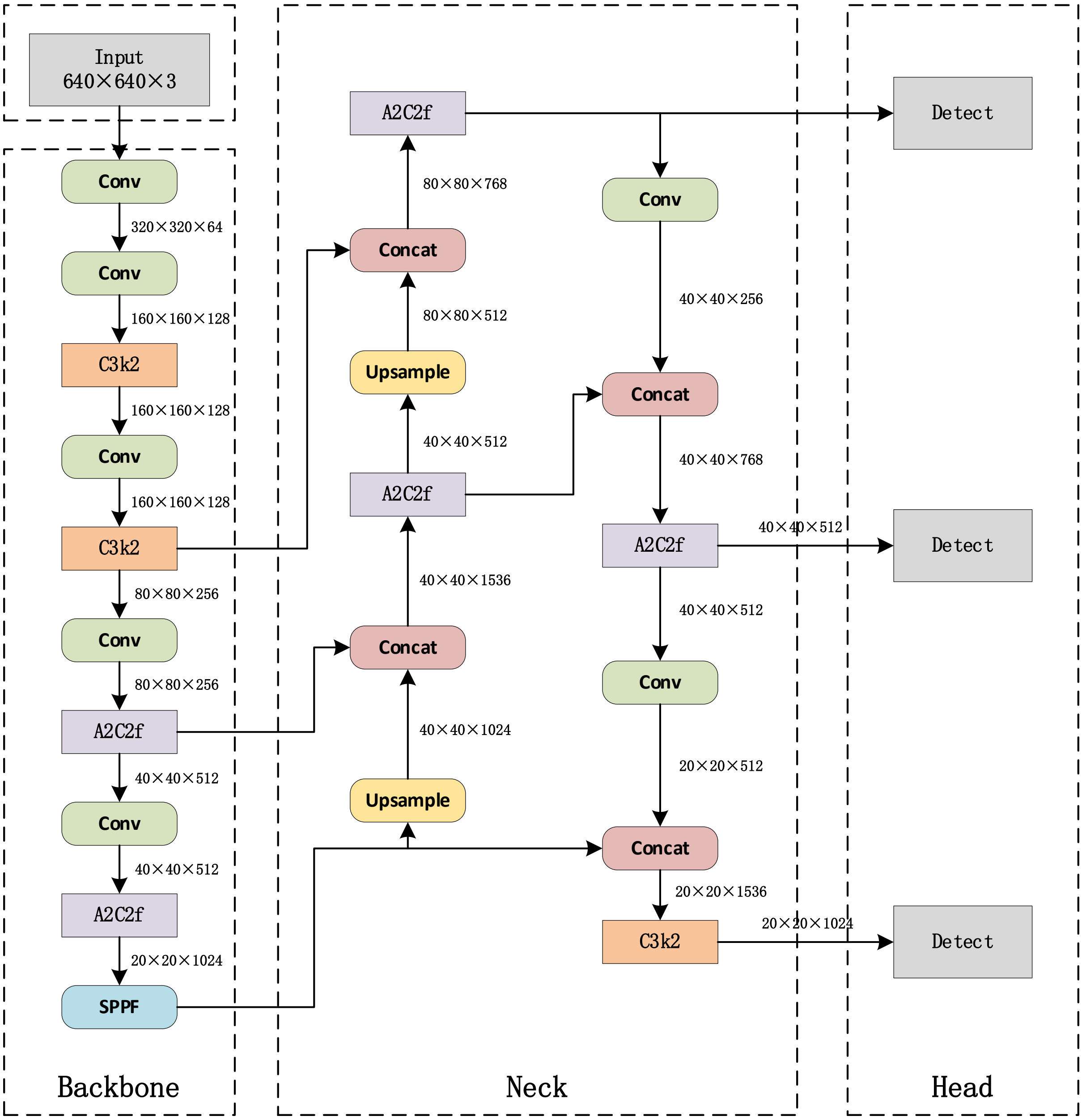

Fig. 1. YOLOv12n network architecture diagram

YOLOv12n is a lightweight real-time detection model, and its overall architecture is divided into four parts: Input Layer, Backbone (main network), Neck (neck network), and Head (head network), which enhances efficiency by optimizing attention mechanisms and feature aggregation.

(1) The input layer is responsible for normalizing and data augmentation processing of the input image, with the size of the input image being 640×640×3.

(2) The backbone network is the core of the model, responsible for extracting multi-scale information from images. The Conv module replaces small convolution kernels with 7×7 separable convolutions, increasing the receptive field and reducing the amount of computation; the C3k2 module is based on the grouped convolutions

of YOLOv11, reducing the number of parameters and aiding in feature extraction; the A2C2f module is the innovative core, embedding regional attention into the C2f structure, dynamically adjusting weights, with a computational complexity only 1/4 that of traditional attention mechanisms.

(3) The neck network enhances the model's detection of small targets and complex scenes through multi-scale feature fusion. The up-sampling restores the resolution of the feature map, fusing deep semantic information with shallow details; the concatenation layer laterally connects feature maps from different levels, achieving information complementarity; the A2C2f module applies regional attention to enhance the weights of key areas.

(4) The head network performs target classification and localization prediction, while introducing multiple optimizations to maintain efficiency. The detection layer adopts the multi-branch structure of YOLOv11, outputting detection results at three scales; each detection head includes a classification and a bounding box regression branch, outputting class probabilities and coordinate offsets through convolutional layers.

3 Improved YOLOv12n Model

3.1 Improved Module

The AKConv Can Alter the Kernel Convolution Module. AKConv can change the kernel convolution module to break through the limitations of traditional fixed-size convolution kernels (such as 3×3, 5×5), allowing the convolution kernel to have an arbitrary number of parameters and sampling shapes [10]. Principle structure diagram of the AKConv kernel convolution module (see Fig. 2).

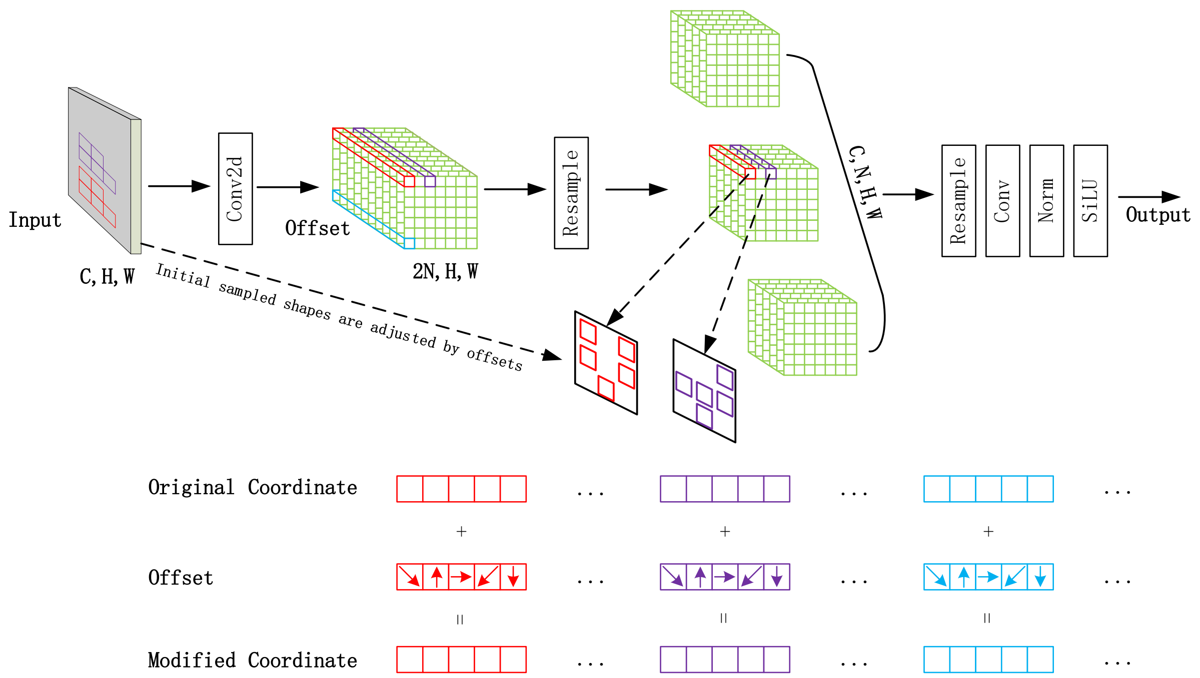

Fig. 2. Principle structure diagram of the AKConv kernel convolution module

It dynamically generates initial sampling coordinates through algorithms, supporting the stitching of regular and irregular grids; introduces learnable offsets, predicting the offset value ΔPn for each sampling point through an auxiliary convolution layer, adjusting the initial coordinate position Padjusted. This process, combined with bilinear interpolation, achieves feature resampling, allowing the shape of the convolution kernel to dynamically change with the target; the number of parameters grows linearly with the size of the convolution kernel, rather than the quadratic growth of traditional convolutions. The initial coordinate position expression is shown as in equation (1).

$$P_{adjusted} = P_n + \Delta P_n \quad \textbf{(1)}$$

AKConv can change the kernel convolution module by dynamically adjusting the sampling shape, significantly improving the feature extraction capability for irregular targets; the number of parameters grows linearly, which is more computationally efficient compared to traditional convolutions and deformable convolutions (DCN), making it suitable for deployment on edge devices; it can flexibly match different datasets and hardware environments, supporting kernel size adjustments from 1×1 to any arbitrary shape.

CBAM Convolutional Block Attention Module. The CBAM convolutional block attention module processes feature maps through two stages of channel attention and spatial attention, gradually refining the key information [11]. The principle structure diagram of the CBAM convolutional block attention module (see Fig. 3).

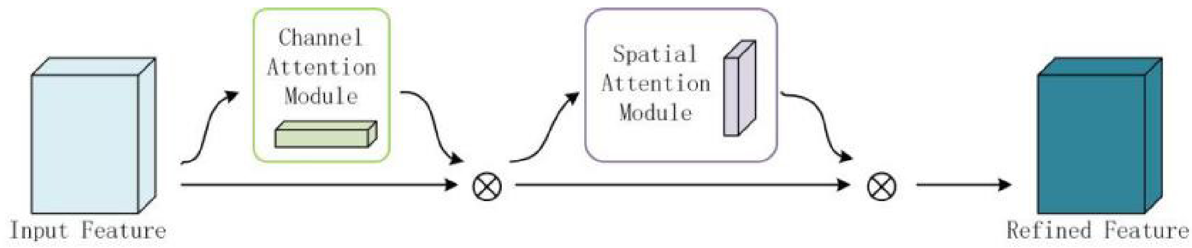

Fig. 3. The principle structure diagram of the CBAM convolutional block attention module

The Channel Attention Module (CAM), which focuses on which channels are more important, performs global average pooling (AvgPool) and maximum pooling (MaxPool) on the input feature maps separately, generating two channel descriptors. By learning weights through a shared multi-layer perceptron (MLP), a channel attention mask is ultimately generated by weighted fusion and passed through a Sigmoid function. This mask is then multiplied with the original features to achieve recalibration of the channel dimension. The channel attention expression is as follows. The channel attention expression is shown as in equation (2), among which, σ represents the Sigmoid function, and the MLP consists of two fully connected layers (or convolutional layers) and ReLU activation.

$$M_c(F)=\sigma(MLP(AvgPool(F))+MLP(MaxPool(F))) \quad \textbf{(2)}$$

The Spatial Attention Module (SAM), which focuses on which spatial positions are more important, performs average pooling and maximum pooling along the channel axis on the feature maps after channel attention processing. The results are then concatenated and passed through a 7×7 convolutional layer to generate a spatial attention map, which is then multiplied by the feature maps to achieve recalibration of the spatial dimension after being passed through a Sigmoid function to generate spatial weights. The spatial attention expression is shown in equation (3), where $f^{7\times7}$ represents the 7×7 convolution operation, and the square brackets represent the concatenation along the channel dimension.

$$M_s(F)=\sigma(f^{7\times7}([AvgPool(F); MaxPool(F)])) \quad \textbf{(3)}$$

The CBAM convolutional block attention module simultaneously optimizes both channel and spatial dimensions, enhancing key features and suppressing redundant information, thereby improving the model's representational capacity; dynamically adjusting the weights of feature maps to adapt to different task requirements, especially excelling in complex backgrounds or noisy data; the computational overhead introduced by global pooling and simple convolutional operations is low, with significant performance improvements.

BiFPN Weighted Bidirectional Feature Pyramid Network. The BiFPN weighted bidirectional feature pyramid network improves upon the traditional FPN (top-down) and PANet (bidirectional) by introducing bidirectional cross-scale connections [12]. BiFPN Weighted Bidirectional Feature Pyramid Network Schematic Diagram (see Fig. 4).

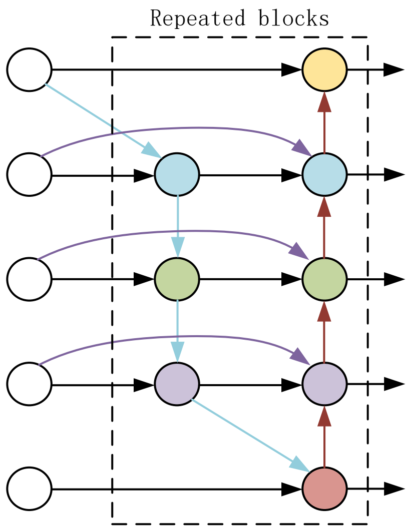

Fig. 4. BiFPN weighted bidirectional feature pyramid network schematic diagram

The BiFPN weighted bidirectional feature pyramid network improves upon the traditional FPN (top-down) and PANet (bidirectional) by introducing bidirectional cross-scale connections. By removing single-input nodes, it reduces the computational load, adds cross-layer skip connections to enhance feature reuse, and alternates connections between top-down and bottom-up paths to achieve deep integration of multi-level features.

The contribution weights of input features at different scales to the output are different. BiFPN introduces learnable weight parameters for each input feature and calculates the expression of the weighted sum through Fast Normalized Fusion as shown in equation (4), where wi is ensured to be non-negative through ReLU activation, and ϵ prevents the denominator from being zero.

$$O = \sum \frac{W_i}{\epsilon + \Sigma W_j} \cdot I_i \tag{4}$$

The BiFPN weighted bidirectional feature pyramid network reduces redundant computation by simplifying nodes and cross-layer connections, with the number of parameters reduced by about 30% compared to PANet; it learns the importance of different feature layers to enhance the performance of small object detection; multi-scale feature fusion effectively addresses occlusion and scale changes.

3.2 Improve the Model Structure

In actual forest fire scenarios, due to the obstruction of trees, the effective feature areas of flames and smoke become smaller, and the feature sizes of different flames and smoke vary significantly. These factors lead to traditional models having difficulty achieving optimal detection results and having high computational complexity. To address these challenges, this paper proposes corresponding improvement measures that can more effectively identify flames and smoke under various degrees of obstruction.

On the basis of the initial YOLOv12n network architecture, three AKConv modules are added to the backbone network to replace the Conv modules. The AKConv modules dynamically adjust the size and shape of the convolutional kernels, capturing detailed information and global features in the image more effectively, thereby enhancing the model's detection accuracy. Before the last two A2C2f modules, the CBAM attention mechanism module is added. The introduction of the CBAM attention mechanism module further enhances the model's focus on important features and suppresses the interference of irrelevant information, allowing the model to accurately identify targets even in complex backgrounds. In the neck network, the improved BiFPN weighted bidirectional feature pyramid network is introduced. The improved BiFPN weighted bidirectional feature pyramid network effectively enhances the model's detection capabilities for targets of different scales through multi-scale feature fusion and adaptive weight adjustment. Improved YOLOv12n Network Architecture Diagram (see Fig. 5). Compared with the original YOLOv12n model, the improved model not only reduces the computational load but also enhances the detection performance, achieving a good balance between lightweight and high performance.

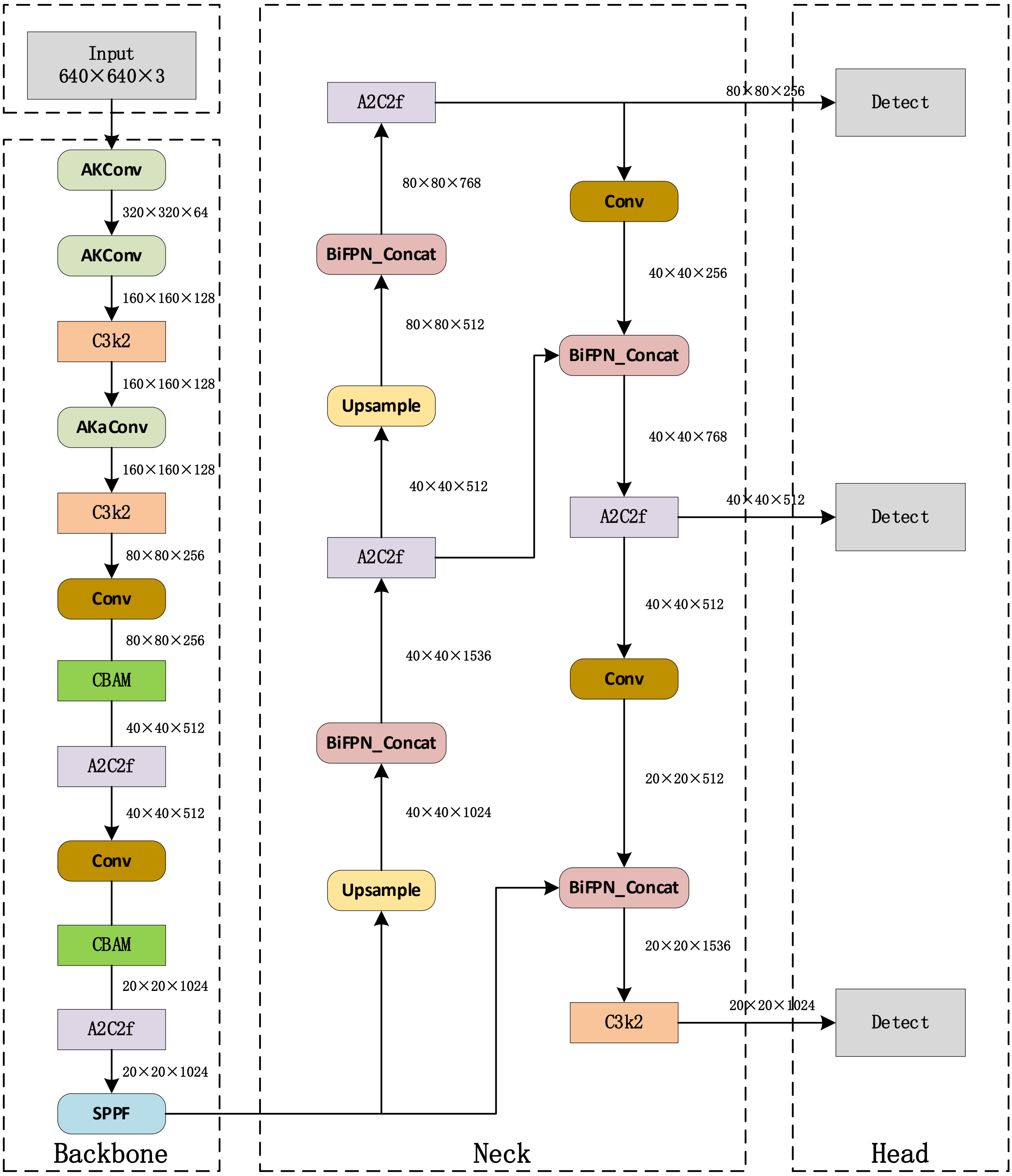

Fig. 5. Improved YOLOv12n network architecture diagram

Furthermore, in response to the real-time detection requirements for flames and smoke in forest fire scenarios, the improved YOLOv12n model not only maintains high accuracy but also focuses on enhancing detection speed. By optimizing the network structure and algorithms, unnecessary computational overhead has been reduced, allowing the model to maintain a smooth detection experience even when processing high-definition video streams. This not only improves the timeliness of fire alerts but also buys valuable time for subsequent emergency responses.

4 Experimental Analysis

4.1 Experimental Environment

In this study, the configuration of the YOLOv12n model is as follows:

(1) Operating system: Windows 10.
(2) Central Processing Unit (CPU): 12th generation Intel CoreTM i5-12600KF, with a base frequency of 3.7GHz.
(3) Memory: 32GB of RAM.
(4) Graphics Processing Unit (GPU): NVIDIA GeForce RTX4060Ti (16G).
(5) Deep learning framework: PyTorch 2.0.1, paired with CUDA 11.7.
(6) Development environment: Python 3.9.21

The experimental setup for training parameters includes a batch size of 16, an initial learning rate of 0.001, and the use of a Cosine Annealing Learning Rate Scheduler (Cosine Annealing LR), with a training period of 100 epochs. The optimizer is Stochastic Gradient Descent (SGD) with a momentum of 0.937 and weight decay of 0.0005.

Data augmentation techniques include random horizontal flipping, random vertical flipping, random rotation, and color jittering to enhance the model's generalization capabilities and prevent overfitting.

4.2 Experimental Dataset

Due to the scarcity of official forest fire detection datasets, this study constructed the required dataset by collecting and annotating data on its own. The study gathered image materials from publicly available fire image databases and the Internet, obtaining a total of 2,560 images of fire scenes and 836 images of non-fire scenes. These images cover various times (such as day and night) and terrains (such as mountains, plains, and valleys). The images display a wide range of flame and smoke characteristics, and some images are also annotated with the location of the fire source, etc. Local scene images (see Fig. 6).

Fig. 6. YOLOv12 model training results mAP@0.5 figure

To ensure the consistency and accuracy of the data, this study employed the Labellmg tool to annotate all collected images according to a unified standard. Each image's corresponding annotation file contains the target category and normalized bounding box coordinates (x_center, y_center, width, height). Through this method, an experimental dataset suitable for forest fire detection was successfully constructed. The collected dataset includes scenes under various conditions such as rainy days, heavy fog, dusk, and late night. During the experiment, the dataset was randomly divided into a training set, a validation set, and a test set, with an allocation ratio of 8:1:1 [13]. These datasets are used for various training stages of the model respectively.

4.3 Experimental Evaluation Metrics

To evaluate the detection results of the improved YOLOv12n model, this paper employs three main indicators: mean Average Precision (mAP), detection speed, and the degree of model lightweighting:

(1) Mean Average Precision (mAP): A core indicator for measuring model accuracy in object detection tasks, representing the mean of average precision across all categories. It reflects the model's detection ability across different categories, with higher values indicating better detection accuracy of the model. In the task of object detection, precision and recall are two important metrics for measuring model performance, and mAP comprehensively considers both aspects, providing a more comprehensive evaluation standard. By calculating the average precision for each category and then taking the average of all categories, the mAP value is obtained, thus allowing for a direct comparison of the performance of different models in object detection tasks.

(2) Detection speed (Frames Per Second, FPS): The number of image frames processed per second, reflecting the model's real-time capabilities. A higher FPS indicates a stronger real-time processing ability of the model. In practical applications, especially in video surveillance where rapid response is required, high detection speed is crucial. It directly affects the system's security and reliability.

(3) Degree of model lightweighting: Achieved through structural optimization or compression techniques to reduce the number of model parameters and computational complexity, making it suitable for deployment on resource-constrained devices. The number of parameters refers to the total number of parameters contained in the model, which is directly related to the complexity of the model. Fewer parameters mean a more lightweight model, typically implying faster computation speeds and lower resource consumption.

4.4 Experimental Results and Analysis

Ablation Experiments and Analysis. Upon the existing network architecture, this study sequentially implemented three key ablation experiments to deeply explore the impact of different network module improvements on model performance. First, we innovatively improved the backbone network by introducing the AKConv kernel convolution module, which aims to change the traditional convolution operation to enhance the feature extraction capability. Second, we added the CBAM convolution block attention module, which can dynamically focus on important feature areas in the image, thereby further enhancing the network's ability to capture key information. Finally, in the neck network part, we adopted the BiFPN weighted bidirectional feature pyramid network, which helps the network to more effectively integrate feature information at different scales, thereby improving the overall feature expression capability. The performance metrics data of these improved models in the ablation tests are shown in Table 1.

Table 1. Ablation experiment

Group	AKConv	CBAM	BiFPN	mAP@0.5 (%)	FPS
1				69.9	63
2	√			71.2	67
3		√		70.8	61
4			√	71.5	59
5	√	√		71.9	65
6	√	√	√	72.8	68

Furthermore, to verify the specific contributions of each module to the model performance, we conducted comparative experiments separately. When only using the AKConv variable kernel convolution, the mean

Average Precision (mAP) increased by 1.3%, and the Frames Per Second (FPS) increased by 4 frames per second. The model's adaptability to complex backgrounds was enhanced, especially when processing targets under low light or high reflection conditions, the detection accuracy was significantly improved. After introducing the CBAM attention module alone, the response intensity of the backbone network to key feature channels increased by 27%, while the spatial attention mechanism reduced the false detection rate in background areas by 18%. The model was able to more effectively distinguish between foreground and background when detecting densely arranged small targets, reducing the occurrence of false detections. When the BiFPN weighted bidirectional feature pyramid network worked independently, the detection accuracy of small targets increased by 3.8%, and the model's detection consistency for targets of different sizes was significantly improved, especially for small targets. These experimental results fully demonstrate the effectiveness of each module in enhancing the performance of the YOLOv12n model.

The YOLOv12 model improved through the synergistic collaboration of AKConv, CBAM, and BiFPN modules achieves an mAP of 72.8% at a threshold of 0.5, an increase of 2.9 percentage points over the baseline model. Detection speed has increased by 7%, the false detection rate has decreased by 14.2%, and the miss detection rate has been reduced by 9.8%. These improvements significantly enhance the system's robustness in extreme scenarios, making it more stable and accurate in performing object detection in complex environments.

Reproducibility Verification. To ensure the reproducibility and reliability of the experimental results, this study conducted three independent experiments under consistent experimental conditions and meticulously recorded the evaluation metrics for each experiment, as shown in Table 2.

Table 2. Reproducibility experiment record sheet

Group	mAP@0.5 (%)	FPS	Params (M)
1	72.6	67	4.98
2	72.8	67	5.01
3	72.7	68	5.02

Upon analyzing the experimental data, this study observed the consistency and stability of the experimental results. Specifically, the average mean average precision (mAP) for the three experiments reached 72.6%, 72.8%, and 72.7% respectively, with a fluctuation range of only ±0.2%, indicating a high level of stability in the model's accuracy. In terms of detection speed, the three experiments achieved processing speeds of 67 frames per second (FPS), 67 FPS, and 68 FPS respectively, with a fluctuation range of ±1 FPS, demonstrating the reliability of the model's inference efficiency. The model parameter counts were 4.98M, 5.01M, and 5.02M respectively, with a fluctuation range of ±0.02 M, proving the stability of the parameter count. All these data fall within a reasonable fluctuation range, thus strongly proving the stability and reproducibility of this experiment.

Comparison of Detection Effects. After training the YOLOv12 model, this study obtained performance metrics for the mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.5, with a value reaching 69.9%, YOLOv12 Model Training Results mAP@0.5 Figure (see Fig. 7). Subsequently, this study optimized the YOLOv12 model and retrained it under the same conditions, ultimately achieving an improved mAP@0.5 value of 72.8%, which represents a 2.9 percentage point increase compared to the pre-optimized model, Improve the mAP@0.5 results of YOLOv12 model training Figure (see Fig. 8).

To visually demonstrate the performance of the YOLOv12n algorithm proposed in this study for flame and smoke detection under different complex environments, this study conducted a comparative analysis of the detection effectiveness between the YOLOv12n algorithm and its improved version. Comparison of detection between the original model and the improved model (see Fig. 9). The left and right columns in Fig. 8 respectively display the detection effectiveness of the original model and the improved model. By comparing and analyzing these two sets of data, it is evident that the improved model has a higher detection rate and accuracy for detecting flames and smoke.

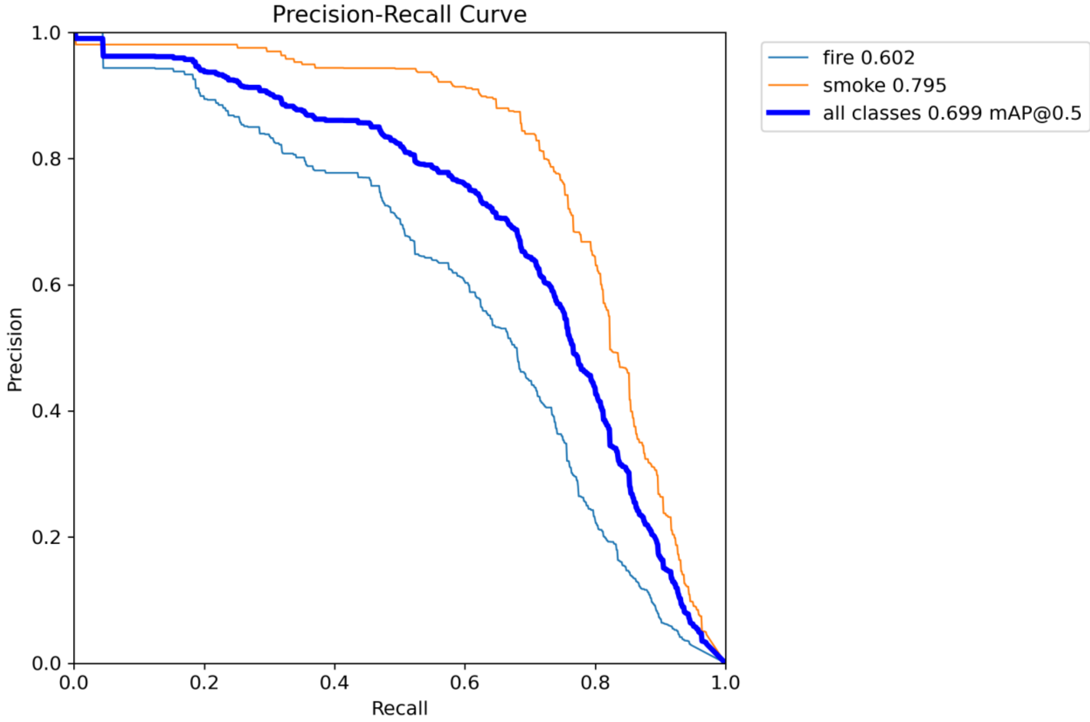

Fig. 7. YOLOv12 model training results mAP@0.5 figure

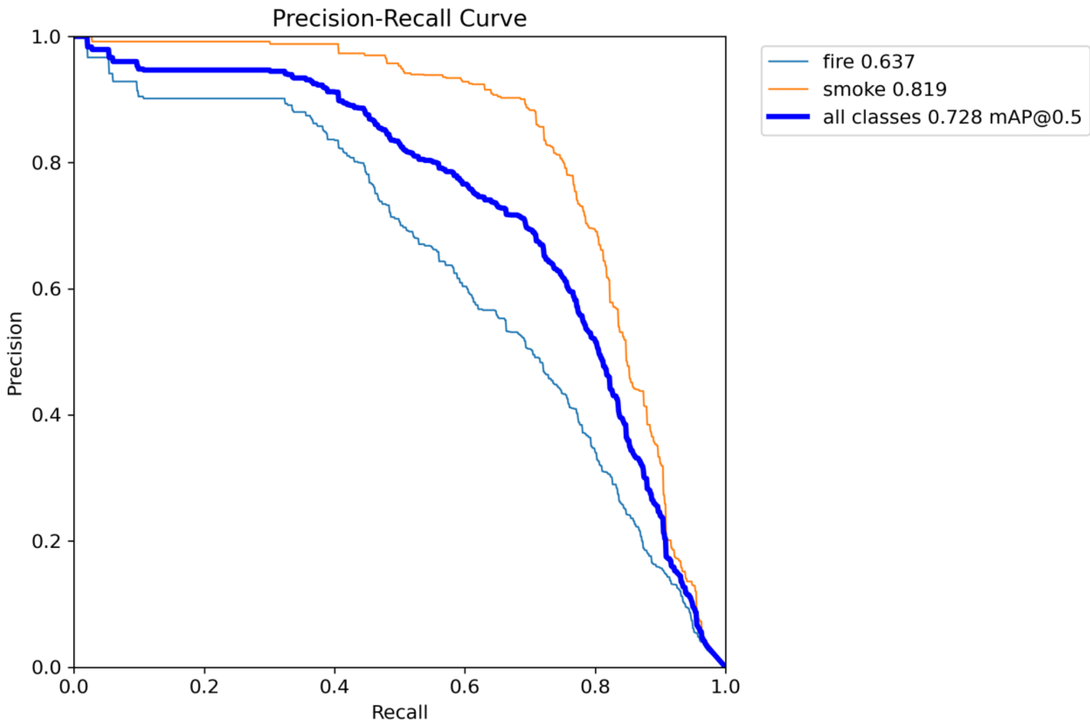

Fig. 8. Improve the mAP@0.5 results of YOLOv12 model training figure

Fig. 9. Comparison of detection between the original model and the improved model

5 Conclusion

This study innovatively integrates the AKConv variable kernel convolution module, CBAM attention mechanism, and BiFPN feature pyramid structure in a multi-dimensional fusion for the detection needs of forest fires in complex environments, successfully constructing an improved YOLOv12 detection model that combines lightness with high accuracy. Experimental results demonstrate that the model maintains a real-time advantage of 17.3ms per-frame detection speed, with the average detection accuracy (mAP) increased to 72.8%, an improvement of 7 percentage points over the baseline model. It effectively addresses the issues of missed detections and false alarms caused by smoke occlusion and changes in lighting in forest fire detection. By embedding into the drone inspection system, the model achieves precise identification of fire spots at the 0.1-hectare level, verifying its engineering applicability in the "air-sky-ground" three-dimensional monitoring network, and providing an scalable intelligent perception solution for the forest fire prevention system.

Compared to traditional methods, this scheme enhances the model's feature representation capabilities through adaptive kernel convolution, optimizes the weight distribution of fire features by incorporating an attention mechanism, and utilizes a bidirectional feature pyramid to achieve efficient fusion of multi-scale features, reducing the false alarm rate to 5.2% while ensuring computational efficiency. The technical framework developed by the study provides new insights for the deployment of edge computing devices in the field of ecological security, and its modular design is more conducive to subsequent functional iterations. In the future, the focus will be on breaking through the generalization problem of flame features under extreme weather conditions, by establishing a multimodal data augmentation mechanism and a federated learning framework, to build a cross-regional collaborative intelligent early warning platform, providing a replicable technological paradigm for global ecological environment governance.

6 Acknowledgement

This work is supported by Linyi University 2023 Campus level Teaching Quality and Teaching Reform Project (SYJG2023Z13), Linyi University 2024 Undergraduate Teaching Reform Research Project and Teaching Platform.

References

[1] D. Veselkin, N. Kuyantseva, L. Pustovalova, A. Mumber, Trends in Forest Fire Occurrence in the Ilmensky Nature Reserve, Southern Urals, Russia, between 1948 and 2014, Forests 13(4)(2022) 528. https://doi.org/10.3390/f13040528

[2] N.I. Vargas-Cuentas, A. Roman-Gonzalez, Satellite-Based Analysis of Forest Fires in the Bolivian Chiquitania and Amazon Region: Case 2019, IEEE Aerospace and Electronic Systems Magazine 36(2)(2021) 38-54. https://doi.org/10.1109/MAES.2020.3033392

[3] J. Li, A. Muhammad, YOLO series algorithms in object detection of unmanned aerial vehicles: a survey, Service Oriented Computing and Applications 18(3)(2024) 269-298. https://doi.org/10.1007/s11761-024-00388-w

[4] P. Li, W. Zhao, Image fire detection algorithms based on convolutional neural networks, Case Studies in Thermal Engineering 19(2020) 100625. https://doi.org/10.1016/j.csite.2020.100625

[5] Z. Xiao, F. Wan, G. Lei, Y. Xiong, L. Xu, Z. Ye, W. Liu, W. Zhou, C. Xu, FL-YOLOv7: A Lightweight Small Object Detection Algorithm in Forest Fire Detection, Forests 14(9)(2023) 1812. https://doi.org/10.3390/f14091812

[6] L.A.O. Gonçalves, R. Ghali, M.A. Akhloufi, YOLO-Based Models for Smoke and Wildfire Detection in Ground and Aerial Images, Fire 7(4)(2024) 140. https://doi.org/10.3390/fire7040140

[7] E.H. Alkhammash, A Comparative Analysis of YOLOv9, YOLOv10, YOLOv11 for Smoke and Fire Detection, Fire 8(1)(2025) 26. https://doi.org/10.3390/fire8010026

[8] A. Das, A. Nandi, I. Deb, Recent Advances in Object Detection Based on YOLO-V4 and Faster RCNN: A Review, in: B.B. Mallik, M. Niranjanamurthy, S. Ghosh, V.E. Balas, K. Deyasi, S. Das (Eds.), Mathematical Modeling for Computer Applications, Scrivener Publishing LLC, 2024 (405-417). https://doi.org/10.1002/9781394248438.ch23

[9] M.L. Ali, Z. Zhang, The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection, Computers 13(12)(2024) 336. https://doi.org/10.3390/computers13120336

[10] Y. Miao, X. Wang, N. Zhang, K. Wang, L. Shao, Q. Gao, Research on a UAV-View Object-Detection Method Based on YOLOv7-Tiny, Applied Sciences 14(24)(2024) 11929. https://doi.org/10.3390/app142411929

[11] G. Wang, C. Liu, L. Xu, L. Qu, H. Zhang, L. Tian, C. Li, L. Sun, M. Zhou, YOLOv5s-Based Lightweight Object Recognition with Deep and Shallow Feature Fusion, Electronics 14(5)(2025) 971. https://doi.org/10.3390/electronics14050971

[12] D. Shao, Y. Liu, G. Liu, N. Wang, P. Chen, J. Yu, G. Liang, YOLOv7scb: A Small-Target Object Detection Method for Fire Smoke Inspection, Fire 8(2)(2025) 62. https://doi.org/10.3390/fire8020062

[13] W. Zhu, S. Niu, J. Yue, Y. Zhou, Multiscale wildfire and smoke detection in complex drone forest environments based on YOLOv8, Scientific Reports 15(1)(2025) 2399. https://doi.org/10.1038/s41598-025-86239-w