Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 13, No. 1, 2025

Large-scale Fire Detection based on YOLOv8 Lightweight Model and its Improved System

Rui Wen *

College of Computer and Information Science, College of Software, Southwest University, Chongqing, 400715, China
*** Corresponding author Email:** wr18223295633@email.swu.edu.cn

Abstract: In recent years, global warming, population increase, the expansion of human production and living areas and many other factors have led to the increasing frequency of forest fires, forest fire prevention and response has become more and more important. Nowadays, computer vision methods and video surveillance cameras have a wide range of applications, has shifted from traditional detection methods to machine deep learning technology, these technologies have a very important role in promoting and significance of fire detection, image detection can be extracted through the dataset training after the object features, compared with the traditional fire detection methods, it can detect and warn of fires in a short period of time, and the coverage of the area is wider, and can realize a wide range of detection. It can realize a wide range of detection. YOLOv8 and its improved models have great advantages in this regard. The application of YOLOv8-based model in fire detection can greatly reduce the consumption of manpower and financial resources for detection, and reduce the economic losses caused by fire. This article mainly tested yolov8 with Fire smoke detection and compared it with convolutional neural network and multimodal model as well as based on yolov8 model by adding modification module to achieve light weight, wide range detection and other functions.

Keywords: Forest Fire Detection; Multimodal Fusion; Wide Range Monitoring; Lightweight System; Low Latency Feedback.

1. Introduction

Fire is a major natural disaster that occurs frequently around the globe, posing a serious threat to human society, the economy and the ecological environment. According to statistics, the annual direct economic losses caused by fires worldwide exceed hundreds of billions of dollars and lead to tens of thousands of casualties, as shown in Table 1, the loss data caused by fires from 2012 to 2020 shows that it poses a huge threat to society and individuals.

Table 1. Consolidated table of data from the China Statistical Yearbook: Forest fires nationwide (2012-2020 data)

Year	Number of forest fires (times)	Total fire area (ha)	Area of victimized forests (ha)	Number of casualties	Depreciation of other losses (in ten thousand yuan)
2012	3966	2397	1568	1	0
2013	3929	2347	1582	0	0
2014	3703	2080	1620	2	1
2015	2936	1676	1254	6	0
2016	2034	1340	693	1	0
2017	3223	2258	958	4	3
2018	2478	1579	894	3	2
2019	2345	1534	802	8	1
2020	1153	722	424	7	0

Fig 1. Forest fire detection device

Traditional fire detection technology mainly relies on smoke sensors, temperature sensors and manual monitoring, which has problems such as high response delay, high false alarm rate and limited coverage. With the acceleration of urbanization and the increase of complex building structures, the demand for real-time, accurate and robust fire detection is becoming increasingly urgent. In this context, fire detection methods based on computer vision and deep learning techniques have gradually become a research hotspot, which can realize non-contact and high-precision fire recognition and early warning by analyzing video, image and multimodal data. As shown in Figure 1, this is a forest fire detection device used for large-scale detection.

Traditional fire detection technologies mainly rely on physical sensors and manual monitoring, including smoke sensors, temperature sensors, and UV/IR flame detectors. These methods trigger alarms by capturing specific physical signals, such as smoke particle concentration, temperature surge, flame spectrum, etc., with low cost, simple deployment and other advantages, but its limitations are significant: a single dimension of information, relying only on a single signal, can not be integrated to determine the characteristics of the fire source; response passive and delayed, need to wait

for the physical parameters to reach the threshold, the negative ignition or early fire detection lag; high false alarm rate, susceptible to environmental interference; limited coverage, the sensor needs to wait for the physical parameters to reach the threshold, for the negative ignition of fire or early fire Detection lag; high false alarm rate, susceptible to environmental interference; limited coverage, the sensor needs to be in close contact with the fire source, it is difficult to apply to open space or complex scenarios. In addition, although manual monitoring can make empirical judgments, there are problems such as high labor costs, poor real-time performance, and large subjective errors. The traditional method is essentially a "passive response" mechanism, which lacks the multi-dimensional perception of the dynamic characteristics of the fire and the ability to analyze intelligently.

With the breakthrough of computer vision and deep learning technology, fire image detection technology based on large models has gradually become a research focus. Early research focused on flame color model and motion feature extraction, but the generalization ability of artificially designed features is insufficient in complex lighting and fire-like interfering object scenes. Compared to deep learning-based image detection techniques, there is a generation gap between traditional fire detection methods in terms of sensing dimension, scene adaptability, and functionality expansion. First of all, traditional sensors can only provide the "presence or absence" judgment of fire, while the large model can extract multi-level features such as flame morphology, smoke diffusion dynamics, etc. from the image to realize the localization of the fire source, fire grading and spread prediction. Secondly, the traditional methods in complex environments with high risk of failure, such as haze obstruction, insufficient light at night, and multimodal model through the fusion of visible light, infrared thermal imaging and sensor data, breaking the perception bottleneck of a single signal source, significantly improving the robustness. In addition, the traditional technology is difficult to cover a wide range of scenes, while the UAV equipped with lightweight YOLO model can carry out aerial inspection, and the detection efficiency is improved by tens of times. What's more, the false alarm rate of traditional methods is generally higher than 10%, while deep learning-based target detection algorithms can control the false alarm rate below 5% through end-to-end feature learning, while supporting early fire warning. These gaps highlight the inevitability of big model technology to promote the transformation of fire detection from "single threshold alarm" to "intelligent active prevention and control".

In recent years, the development of machine learning and deep learning has completely changed the field of fire detection. The model based on Convolutional Neural Network (CNN) performs very well in image classification and target detection tasks, while the single-stage target detection framework YOLO (You Only Look Once) has become an important tool for fire detection in dynamic scenarios by virtue of its high efficiency and real-time performance. Meanwhile, the deep learning method based on multi-source information fusion significantly improves the detection accuracy and environmental adaptability of fire monitoring systems by integrating multimodal data such as visual images, infrared thermal imaging, temperature sensing, and smoke concentration. This type of advanced model not only effectively solves the inherent defects of traditional detection technology but also opens innovative technological paths for intelligent fire warning in complex environments.

CNN, multimodal models and YOLO series models each have their own advantages and limitations in fire detection. CNN is mainly used for tasks such as static image analysis, extracts flame features through multilayer convolution, and performs well in smoke recognition, etc. However, it has high computational complexity, is weak in modeling dynamic flames and is prone to overfitting, but it is still the basic architecture for fire feature extraction. Multi-modal model fusion of multi-source data to improve the robustness of fire detection in extreme environments, using cross-modal information complementary to reduce the false alarm rate, however, computationally complex, there are spatial and temporal alignment, high cost of annotation and high hardware resource requirements, the future focus on lightweight fusion and self-supervised learning. YOLO series models to single-stage detection framework of high efficiency, suitable for real-time video surveillance and mobile deployment, can synchronize the It can detect multiple types of targets and support real-time processing, this article used the dataset of Fire smoke detection to train the relevant modules of YOLOv8 and achieved good results, but it has a high leakage detection rate for small-scale flames and sparse smoke, limited generalization ability, and is still affected by target size and data distribution despite improvements.

In summary, the above three provide a theoretical basis for low-latency large-scale fire detection and provide a reference direction for scholars to improve the training of the model when doing related research, and also provide them with a part of the examples of previous scholars as a reference.

2. Related Work

2.1. CNN

Convolutional Neural Network (CNN) is a deep learning model designed for processing grid-like data, the core idea is to extract spatial features through local perception and weight sharing, which mainly contains the following components.

Convolutional Layer: Sliding scanning of input image using convolutional kernel to extract local features such as edges, texture etc. Reduces the number of parameters and preserves spatial information through local concatenation, suitable for image translation invariance.

Pooling Layer: Downsampling the feature map to reduce computational complexity and enhance feature robustness. Compress the feature dimension to prevent overfitting and improve the model's tolerance to small deformations.

Activation Function: Solving the gradient vanishing problem and accelerating convergence.

Fully Connected Layer: Mapping high-level features to classification labels to complete the final decision is usually located at the end of the network, connecting all neurons.

CNN, as a cornerstone model of computer vision, significantly improves detection accuracy and scene adaptation through automatic feature learning and end-to-end optimization in fire detection. However, its limitations in real-time, dynamic modeling and small target detection still need to be further broken through by combining lightweight design, time-series network and multimodal fusion techniques.

In recent years, many scholars have made a lot of improvements based on CNN to improve the performance of detecting fire to some extent. The Khan Muhammad research team used an improved SqueezeNet architecture for model

fine-tuning and developed a fire detection system called CNNFire. This study innovatively proposes an intelligent feature selection algorithm that can automatically identify and extract the most discriminative feature maps for fire areas in convolutional layers. These feature maps filtered by algorithms demonstrate superior accuracy in fire segmentation tasks compared to traditional manually designed features. Based on the segmentation results generated from these feature maps, researchers can deeply analyze the dynamic characteristics of fires, including key parameters such as fire spread speed. In addition, this method can effectively evaluate the development stage and combustion intensity of a fire, providing a more comprehensive analytical dimension for fire monitoring[1]. But there are also motion-based fire detection schemes that may not work properly if there is cloudy and complex weather. And the false positive score of 11.67% is still high and the accuracy needs to be further improved. To achieve high accuracy and low false alarm rate.

Also, Khan Muhammad et al. used a GoogleNet-like model and applied a transfer learning strategy [2]. GoogleNet is an ideal choice due to its excellent classification performance, compact model structure, and good adaptability to memory limited hardware. Through the application of transfer learning technology, researchers have successfully improved the accuracy of flame detection from 88.41% to 94%, achieving significant progress. However, there is still room for improvement in reducing false alarm rates using this method, which points the way for future research.

Barmpoutis P proposed an innovative early fire detection method that combines deep learning and multi-dimensional texture feature analysis techniques based on linear dynamic systems (LDS).This study adopts the Regional Convolutional Neural Network (R-CNN) architecture, which synchronously optimizes target classification and bounding box regression tasks through end-to-end training mechanisms, achieving precise localization and recognition of fire areas[3]. Such an approach significantly reduces the overall computational complexity while improving the performance by significantly reducing the false positive rate due to flame color targets while maintaining a high true positive rate.

The Faisal Saeed research team has developed a multimodal deep neural network architecture that integrates three core model components: a hybrid classifier based on Adaboost algorithm and multi-layer perceptron (MLP) ensemble; Combining Adaboost with Local Binary Pattern (LBP) feature extraction model; Deep Convolutional Neural Networks. To optimize detection performance, researchers specifically used CNN for secondary classification validation of candidate regions (ROIs). This strategy not only significantly reduced the system's false alarm rate, but also improved the overall classification accuracy to an excellent level of 97.8% [4]. However, more datasets are used, the training period is longer, and the mixing of multiple models may lead to a more complex system that is not suitable for popularization.

Although there are a lot of research with more obvious improvements on the basis of CNN, there still exists the problem of misjudging fire detection in dynamic scenes and variable and complex weather. Moreover, the deep network structure of CNN contains a large number of convolutional layers and fully connected layers, with high computational complexity and limited inference speed, even if there is an improvement on the basis of this, it is still difficult to achieve real-time detection, which is likely to lead to the detection of the fire after it spreads to trigger the alarm, and fail to achieve the effect of early warning. And the CNN high-level feature map has a low resolution after many pooling, and the detail information of small targets such as long-distance small fires is seriously lost. In the early stage of the flame, long-distance small fires only account for a few pixels of the image, and it is difficult for CNN to carry out effective recognition.

2.2. Multimodal:

Multimodal deep learning refers to improving the perception and decision-making ability of a model by jointly learning information from multiple heterogeneous data. In fire detection, modalities usually include:

Visual modalities: visible camera images, infrared thermal imaging.

Physical modalities: temperature sensor data, smoke concentration sensor signals.

Environmental modalities: meteorological data such as wind speed, humidity, Geographic Information System (GIS) maps.

Different modalities provide complementary information, and through cross-modal association learning, it breaks through the perceptual limitations of a single data source, allowing for more accurate detection, while supporting multi-task output.

In recent years, several scholars have also conducted research on multimodal deep learning for fire detection. For example, Sharma A et al. collected multimodal datasets using thermal imaging cameras that can detect temperature changes. This study adopts a multimodal data fusion strategy to improve detection performance through collaborative training of thermal imaging visual data and fire sensor data. In terms of visual modality processing, the research team trained multiple improved Convolutional Neural Network (CNN) architectures using image datasets collected by thermal imaging cameras. At the same time, for sensor data streams, we innovatively adopted bidirectional long short-term memory networks (BiLSTM Dense) and densely connected long short-term memory networks (LSTM DenseDenseNet 201) for temporal feature modeling [5]. The experimental results show that this multi-source data fusion method exhibits significant performance advantages: the CNN model maintains a classification accuracy of over 0.99 in visual modality, while the three temporal models of Dense, BiLSTM Dense, and LSTM DenseDenseNet 201 perform better in sensor data processing than traditional BiLSTM dense models, with the highest accuracy reaching 93.39. The system ultimately achieved a verification accuracy of 99.7%, and the verification loss was maintained at an extremely low level, fully demonstrating the effectiveness of this scheme in improving detection accuracy and reducing false alarm rates.

The Bhamra J K research team has developed a multimodal SmokeNet system and its integrated version SmokeNet Ensemble, specifically designed for multi-source monitoring of wildfire smoke. This system innovatively integrates three heterogeneous data sources: satellite remote sensing data, meteorological sensor parameters, and optical camera images. Based on the original SmokeNet architecture, the study systematically evaluated the improvement effect of multi-source information fusion on fire detection accuracy and timeliness by introducing meteorological observation and satellite fire point detection data. The experimental design focuses on analyzing the impact mechanism of new data

dimensions on system performance indicators, providing important basis for the optimization of multimodal fire warning systems [6]. In their experiments, they used three model architectures, SmokeyNet, SmokeyNet Ensemble, and Multimodal SmokeyNet. The fusion of the three models resulted in an average increase in the speed of detecting smoke by 13.6% and a reduction in the standard deviation by 0.32. the results are summarized in the following table.

The Toulouse T research team has developed an innovative multimodal 3D visual analysis system specifically designed for geometric feature extraction in fire scenes. The system achieves accurate modeling of the three-dimensional spatial characteristics of fires by integrating data collected from multiple 3D vision devices. Based on the registered 3D fire point data, researchers have established a complete method for calculating the geometric features of fires. This method extends the monocular stereo vision technology previously developed by the author. By integrating the three-dimensional point cloud data obtained from the dual vision system, it can accurately reconstruct the spatial characteristics of the fire scene, including key parameters such as the propagation plane parameters of the fire front, spatial position coordinates, spread rate, flame height, burning surface morphology and tilt angle, fire perimeter, burning area, main direction of fire spread, and overall fire volume estimation [7]. Such multimodal fusion is not only able to detect fire effectively, but also able to predict flame propagation, which is very powerful. However, two multispectral cameras (JAI AD-080 GE) are used in the stereo vision system, which makes the cost much higher compared to other large-model detections, and this high cost makes it difficult to detect fires on a large scale.

Chenyu Chaoxia's research team has innovatively developed a flame detection framework that integrates multimodal information. This model uses thermal imaging projection technology to preliminarily locate the flame area and capture the surrounding features of the flame through neighborhood sampling mechanism. The research team has specially designed an attention guidance mechanism based on index matching, using attention weight maps generated from thermal imaging data to enhance the ability to extract regional features of visible light modes. The experimental results show that this multimodal fusion method significantly outperforms the traditional single modal Faster R-CNN model in terms of detection performance [8]. However, the device requires that the positional offsets and orientations of the different modal cameras are not too large and that the flame target is within the field of view of both cameras. Further research is needed for the case where the multimodal images are very poorly aligned or unpaired.

Multi-modal fusion has a very strong performance breakthrough single-modal perception limitations, in haze, night and other scenes, through the infrared and sensor data can also maintain detection capabilities. In fire detection, it can also support multi-task outputs such as fire source localization, fire classification, and spread prediction. However, multimodal parallel processing requires GPU cluster support, difficult to deploy in the edge devices, such as Transformer fusion model training time consuming 3 to 5 times that of a unimodal CNN, and at the same time its application to detect fires will make the cost very high, for example, the camera, you need to label the flame region in the image, the sensor value and its correlation relationship labeling costs, which makes the use of its This makes it very difficult to detect large-scale fires, and the maintenance cost is also very high.

2.3. YOLO

YOLO (You Only Look Once) It is a single-stage target detection algorithm whose core idea is to transform target detection into a regression problem and predict bounding box and category probabilities directly on the image. Compared to two-stage models (e.g., Faster R-CNN), YOLO is known for its high speed and end-to-end optimization, which is especially suitable for fire detection in scenarios with high real-time requirements. In recent years, the iterative upgrades of YOLO have also gradually improved the detection results. YOLOv3/v4: Introduces multi-scale prediction (FPN) and CIoU loss function to improve the detection of small targets in flame and smoke. YOLOv5: adopts PyTorch framework, supports lightweight deployment, and becomes the mainstream choice for fire mobile detection. YOLOv7/v8: optimizes dynamic label assignment and re-parameterized convolution to reduce false alarm rate in complex backgrounds.

Because of the excellent performance of YOLO in real-time detection, a very large number of scholars have conducted research in this area in recent years. More classically, Talaat F M et al. used the Smart Fire Detection (SFD) algorithm to detect fires in real time using computer vision from live cameras or pre-recorded video files in real [9]. The SFD algorithm is an efficient real-time fire detection solution that can quickly respond to potential fire risks. This method adopts the YOLOv8 architecture and can achieve fast and accurate target recognition without relying on regional suggestion networks. Through system optimization, this solution significantly reduces the number of parameters while maintaining its lightweight characteristics, making it particularly suitable for deployment in resource constrained environments. This technology achieves an excellent balance between detection accuracy and computational efficiency. Meanwhile, the model detects fire and smoke with 95.7% and 99.3% accuracy, respectively, and the mean accuracy of detection is 97.5% and 97.5% for both categories, with a high accuracy of 97.1% for all types of fires, which is a very high level of accuracy.

In their study, Goyal S et al. deployed deep learning models on a Raspberry-Pi with neural rods on a drone to help detect fires. They embedded the yolov3 model into a neural bar, and once an image is captured, it is processed and passed to a CNN, which will check for forest fires. And if it detects that the confidence level is exceeded, then it will send a notification [10]. Goyal S et al. fused the YOLO model with CNN and achieved better experimental results with 90% accuracy of YOLO based forest fire detection system. Its sensitivity and specificity are 92% and 90% respectively. This also reflects the powerful plasticity and compatibility of YOLO model.

Wu H et al. proposed a ship fire detection model based on enhanced YOLOv 4-tiny network to address the shortcomings of sensor detection and detection methods of traditional image processing techniques, adding the SE attention mechanism to the enhanced feature extraction network partially generating the I-YOLOv 4-tiny + SENet model [11]. The addition of the SE attention mechanism reduces the dependence on the ship fire dataset, speeds up convergence, and achieves the goal of accurate ship fire detection. The experimental results show that the deep network learning model IYOLOv4 tiny+SE

significantly outperforms other compared models in terms of performance. Compared to YOLOv3 tiny, SSD, YOLOv4 tiny, and I-YOLOv4 tiny, this model has mAP@.5 The indicators have increased by 19.5%, 10.9%, 8.5%, and 2.1% respectively. In terms of accuracy, its advantages reach 16.3%, 10.2%, 7.7%, and 1.9%, while the improvement in recall rate is more significant, reaching 22.1%, 18.1%, 9.2%, and 3.9%, respectively. Although these improvements have brought significant improvements in accuracy, the model still has several limitations. Firstly, its training process relies on a large amount of sample data, and insufficient samples can lead to a decrease in recognition performance. Secondly, training deep convolutional neural networks from scratch takes a long time and requires high computational resources. In addition, high-performance hardware platforms are expensive and require strong computing power and storage capacity support, which poses challenges for the practical application of this model in large-scale detection tasks.

The Cao L research team has made multiple innovative improvements to YOLOv5 architecture. Firstly, they integrated the C3TR module and global attention mechanism into the neck trunk network, significantly improving the visual feature extraction ability of the network receptive field. Secondly, by optimizing the reparameterized convolution module and reconstructing the decoupled head and neck network structure, the convergence speed of model training has been effectively accelerated. In terms of loss function, fully concurrent intersection is used as the bounding box regression loss to achieve optimized configuration of multitask loss function. Of particular note, this study employs a weighted bidirectional feature pyramid network (BiFPN) instead of the traditional feature pyramid structure, enhancing the robustness of multi-scale feature fusion [12]. The experimental results show that the improved model achieves high accuracy, recall, mAP, and mAP@0.50.95 Significant improvements of 4.2%, 3.8%, 4.6%, and 2.2% were achieved in key indicators. However, these performance improvements are accompanied by a decrease of about 5% in FPS, which may have a certain impact on the response speed, missed detection rate, and operational stability of real-time detection systems.

The Bensheng Yun research team has developed a lightweight forest fire detection model FFYOLO, which has undergone multiple innovative improvements based on the YOLOv8 architecture. In terms of feature extraction, researchers have designed a Channel Prior Extended Attention Module (CPDA) to enhance the ability to capture fire and smoke features. The detection head adopts a new hybrid layered structure (MCDH) to improve detection performance. To improve model accuracy, the MPDIoU method was innovatively introduced to optimize bounding box regression and classification performance. The network neck adopts a lightweight GSConv module, which significantly reduces the number of parameters while ensuring accuracy. The use of knowledge refinement strategies during the training process effectively improved the model's generalization ability and reduced false positive rates [13]. The experimental data shows that the improved model achieves 88.8% in mAP0.5 index, which is 3.4% higher than the original model. The number of model parameters is reduced by 25.3%, and the inference speed is improved by 9.3%. Thanks to its lower parameter count and computational complexity, this model is particularly suitable for deployment on edge devices with limited computing resources.

3. Method

If wanting to realize large-scale fire prevention detection, YOLOv8 is a good experimental object, because YOLOv8 is able to integrate the multi-scale attention mechanism and cross-modal lightweight design, considering the high accuracy and real-time, especially suitable for UAVs, satellite remote sensing and other large-scale inspection needs.

Compared to CNN, YOLOv8 supports continuous frame detection of video streams, combined with dynamic label assignment to improve flame trajectory stability, while CNN is mostly used for static image processing, and cannot capture temporal features such as flame diffusion and smoke fluttering. Moreover, traditional CNN requires multiple pooling and fully connected layers, and processing single-frame images is time-consuming, making it difficult to meet the demand for real-time video stream detection. However, YOLOv8 can increase the inference speed to 90-120 FPS to meet the demand for fire warning in a very short time.

Multimodal need to integrate visible light, infrared, sensor multi-modal model, the need for GPU cluster support, difficult to run in real time, and need to align the multi-source data, labeling costs for the single-modal 3-5 times, want to achieve better detection results also need high-cost hardware support, it is difficult to use on a wide scale. YOLOv8 is much more lightweight, does not need so much arithmetic support, and by adding modifications to the module can increase the application of each scenario. YOLOv8 is more lightweight, does not require so much arithmetic support, and by adding and modifying the module to increase the application of each scene, ensuring that the cost is low, to meet the wide range of use.

So why choose the model YOLOv8 in the YOLO series? Compared with previous YOLO versions (YOLOv3/YOLOv5), YOLOv8 adds P2 detection layer, supports detection of tiny targets with resolution $\geq 4\times4$,the small target mAP@0.5 increases to 65% and classification and regression tasks are optimized independently. YOLOv3/YOLOv5 has weak detection of small flames, and high-level feature resolution is insufficient. YOLOv8 adopts the GhostNet lightweight module, the model volume is compressed to 3.8 MB, and it supports the deployment of edge devices, which is convenient for large-scale deployment. YOLOv8 can also improve the robustness of complex lighting and smoke occlusion scenes after introducing multi-scale feature fusion and SENet/CBAM attention mechanism. Compared with the new YOLO model, although the new model has improved in accuracy, the complexity of the model is high and it requires high-performance hardware support, so the cost of wide-scale use will be much higher than YOLOv8.

YOLOv8 also has significant improvements in model architecture, using an enhanced version of CSPDarknet53 on the backbone network to reduce computational redundancy by connecting partially across stages and introducing a deeper network structure to improve feature extraction. The multi-scale feature fusion structure of PAN-FPN (Path Aggregation Network + Feature Pyramid Network) is used on the neck to enhance the combination of shallow details and deep semantic information, and to improve the small target detection effect. The detection head is changed to an anchorless design, which directly predicts the center point and width and height of the target, simplifying the model and reducing the need for hyper-parameter tuning. Based on the YOLOv8 model, this paper designs a method to achieve lightweighting by modifying the

module and network structure and other methods, such as replacing conv with the Ghostconv module.

Overall, YOLOv8 rebalances accuracy and efficiency while maintaining the real-time advantages of the YOLO series through architectural lightweighting and dynamic training strategies. Its performance in the field of fire detection is particularly outstanding, considering the needs of tiny target capture, complex environment adaptability and edge device deployment, and has a greater advantage in fire detection.

Table 2. Performance comparison of YOLOv8

	YOLOv8	YOLOv5
Backbone	CSPDarknet53 Enhanced Deeper Structure	CSPDarknet53
Head	Anchor-Free no anchor	Anchor-BasedBased on predefined anchor frames
Neck	PAN-FPNmultiscale enhancement	PAN-FPN
loss function	Varifocal Loss + CIoU + DFL	Focal Loss + CIoU
COCO mAP	53.9%	50.7%

4. Experimental Design

Because YOLOv8 has a powerful performance, scholars can improve and upgrade it on this basis. Before that, I need to understand the architecture of YOLOv8 (Figure 2).

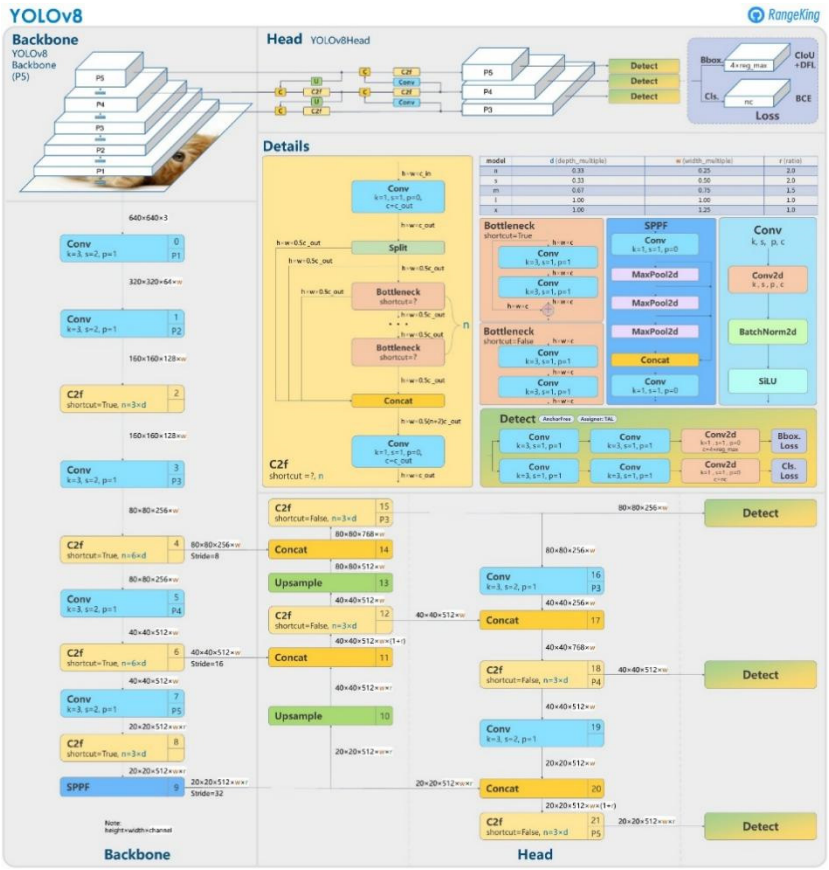

Fig 2. YOLO v8 architecture

As can be seen from Fig. 2, YOLOv8 is mainly divided into two big parts, namely Backbone and Head. in Backbone, the core modules are mainly C2f, conv, and SPPF. the conv module extracts features by convolution operation and integrates the batch normalization and activation functions at the same time, to realize efficient spatial information encoding with the nonlinear feature transformation, which is the basic computational unit of the model.

The C2f module enhances the feature fusion capability while reducing the computation volume through multi-branch gradient flow and cross-stage partial connection, which is the core design for balancing lightweight and performance in YOLOv8.The SPPF module fuses the features of different sensory fields through multi-scale maximal pooling to enhance the robustness of the model to the change of the target size while maintaining a high computational efficiency. Bottleneck. Detect and other modules such as Detect are used for feature compression/expansion, up-sampling to recover resolution, and multi-scale target detection output, respectively, which work together to build the complete inference process of YOLOv8.

The following image shows the result graph of YOLOv8 after thirty rounds of training with 9441 fire and smoke related datasets.

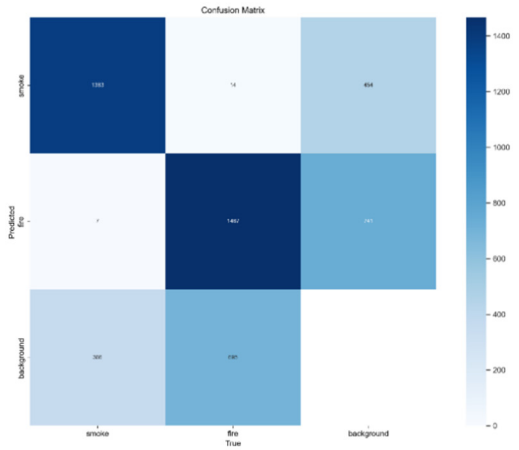

Fig 3. Confusion Matrix

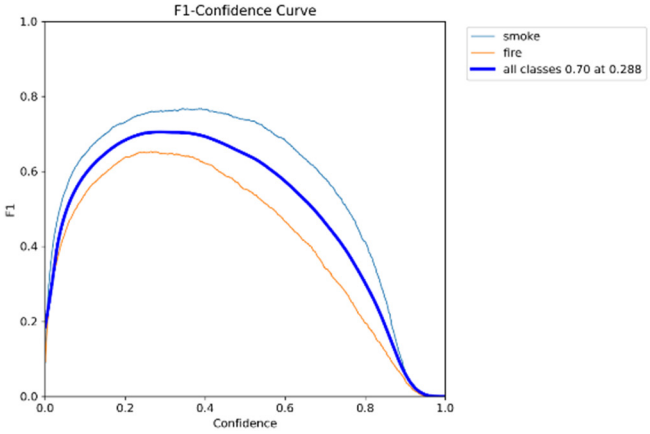

Fig 4. F1-Confidence Curve

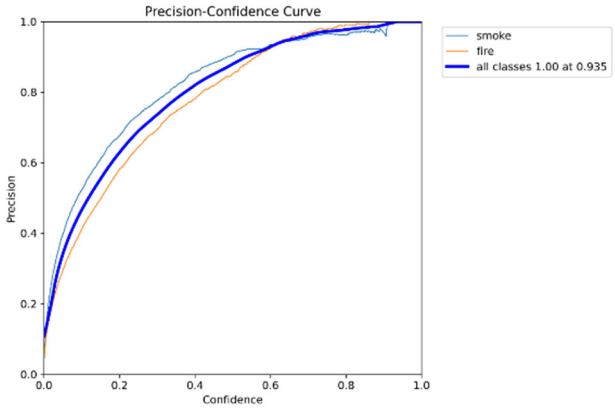

Fig 5. Precision-Confidence Curve

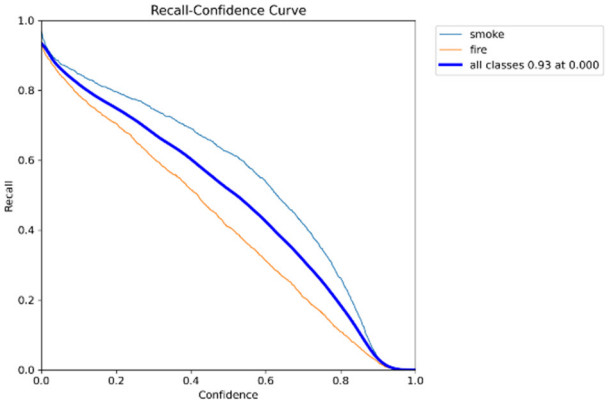

Fig 6. Recall-Confidence Curve

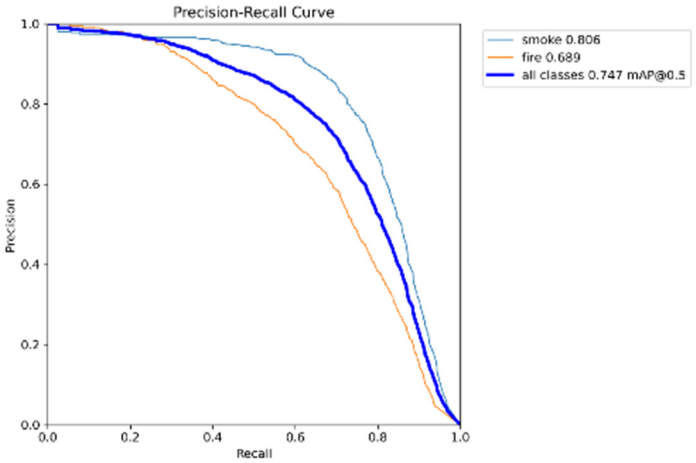

Fig 7. Precision-Recall Curve

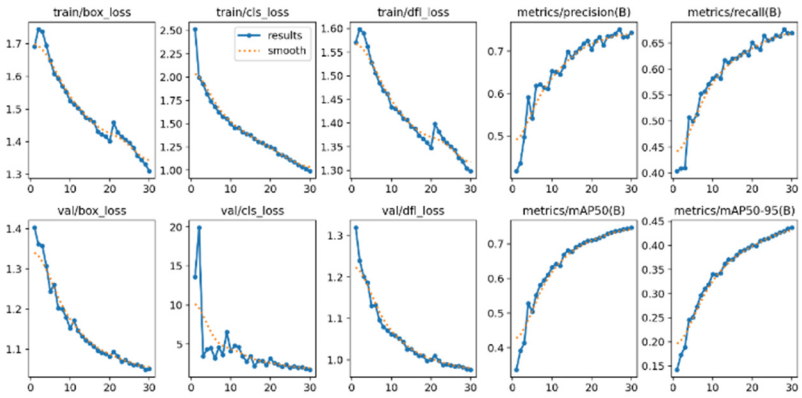

Fig 8. Result

As can be seen from the data presented in Figure 3, the precision rate for detection of SMOKE reaches 78.7% (1383/1756) and the recall rate is 74.7% (1383/1851). Precision for FIRE reaches 67.4% (1467/2176) and recall is 66.2% (1467/2215). Figure 4 shows the F1-Confidence Curve with an F1 value of 0.70 at a confidence threshold of 0.288. Figure 5 shows a precision of 1.0 at a confidence level of 0.935, and Figure 6 shows a recall of 0.93 at a confidence level of 0. The PR Curve's mAP@0.5=0.747 indicates that the model performs well in fire detection. Fig. 7 and Fig. 8 also show comprehensive experimental results.

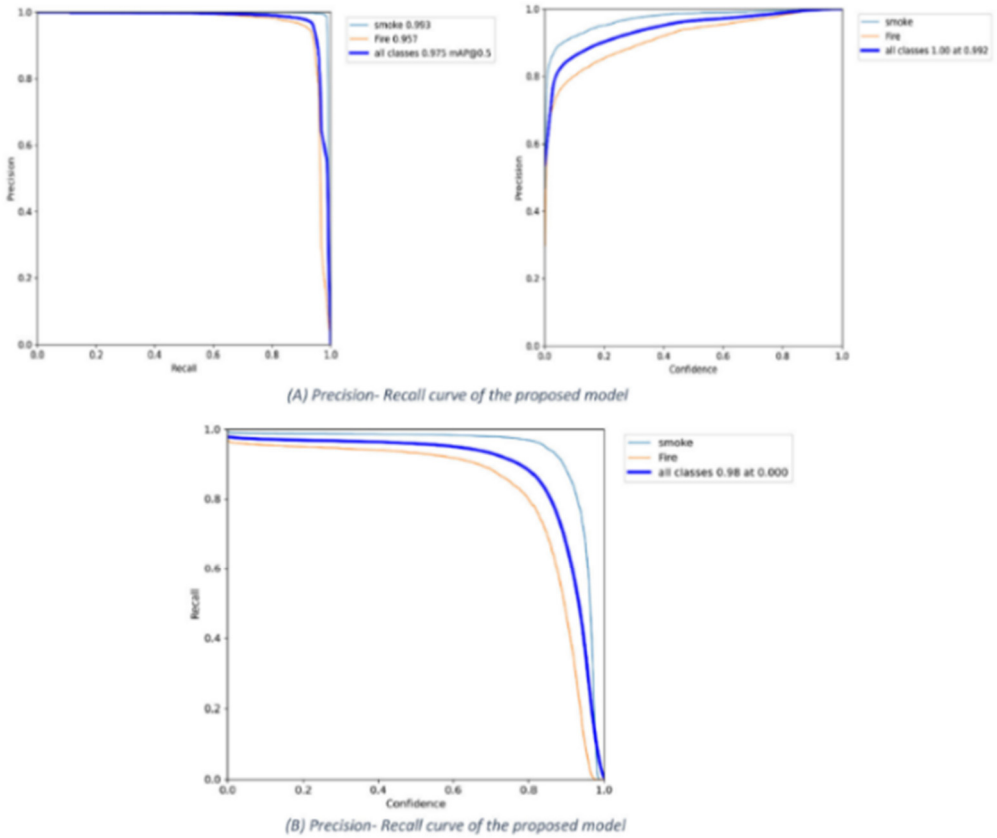

Fig 9. YOLOv8 Improved Model Precision Recall Curves for Smoke and Flame Detection

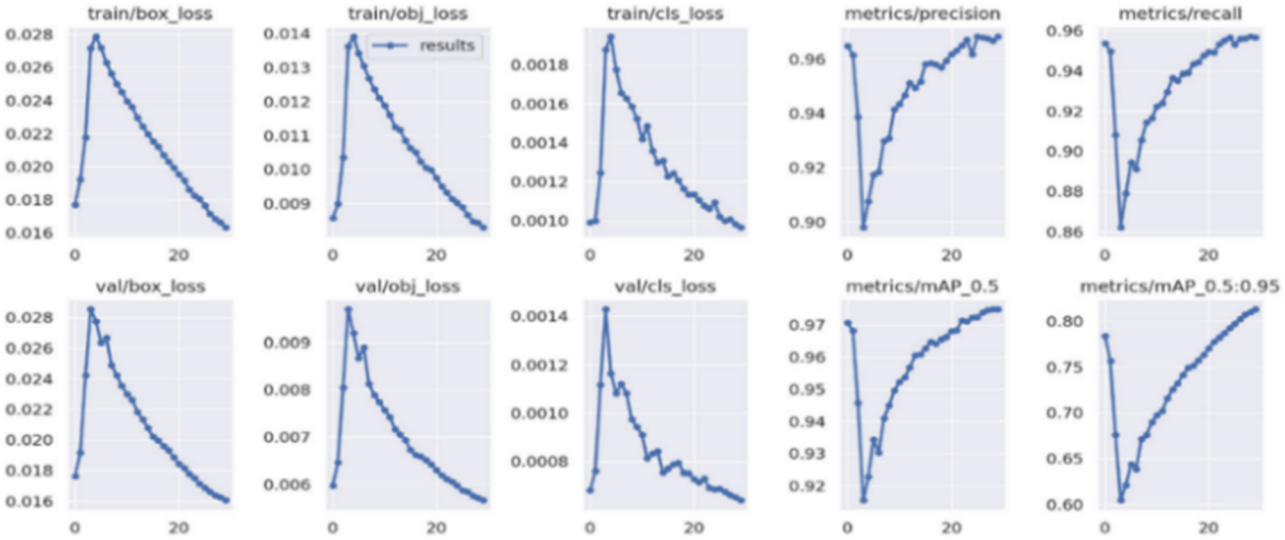

Fig 10. The results of the proposed model

Fig. 9 shows the precision recall curves of the improved model for flame and smoke detection, Fig. 10 shows the overall results of the proposed model in terms of loss, precision and recall, and Fig. 11 shows the different weights of the YOLOv8 model using MCDH. These images show that the improvements made to the YOLOv8 model have a high degree of precision while keeping it lightweight.

YOLOv8 has become the model of choice for real-time target detection due to its efficient inference speed, multi-task compatibility and lightweight design. However, it still has room for improvement in small target detection, complex background robustness and resource consumption. In the

future, the performance can be further improved by strategies such as attention mechanism enhancement, multimodal fusion and quantization compression, such as improving the model from the following aspects:

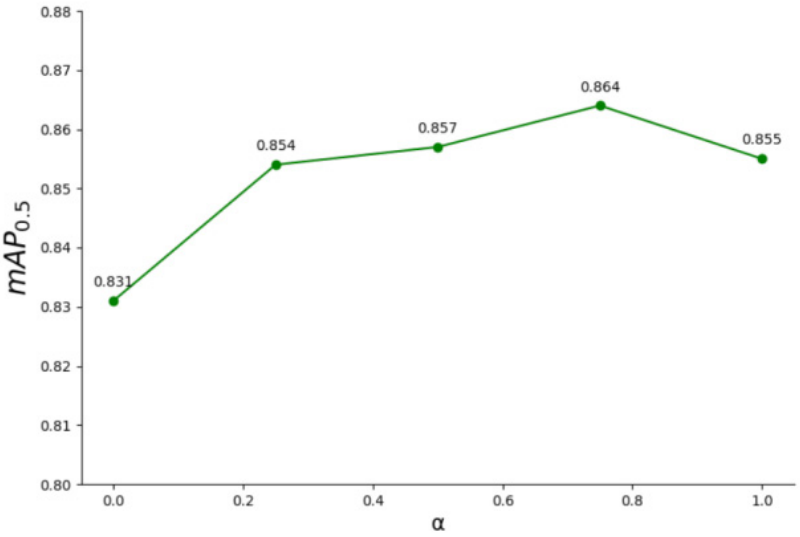

Fig 11. The results of YOLOv8 with different weight coefficients when using MCDH

DEAB module + super-resolution preprocessing can improve in small target detection, The expected boost boosts mAP@0.5 by 10-15%. Improvement of high background false detection rate can be achieved by using CBAM attention + dynamic label assignment, which can reduce more than 20% based on the original false detection rate. To be widely used in fire detection, the model is very necessary to achieve lightweight, in YOLOv8 can be used to the Ghost convolution, adding CBAM module to achieve a reduction in the number of parameters and calculations at the same time to reduce the false detection rate.

5. Conclusion

In the field of fire detection, YOLOv8 is ideal for intelligent early warning systems due to its excellent real-time performance and powerful multi-scale feature extraction capabilities. The model can accurately capture tiny flame features and dynamic smoke targets in early-stage fires, and its lightweight design can be easily deployed on edge devices such as drones and surveillance cameras to achieve millisecond response speed. Compared to traditional detection methods, YOLOv8 performs well in complex environments, effectively overcoming light variations and dust interference, and significantly reducing the false alarm rate while ensuring high accuracy.YOLOv8's technological advantage stems from three major innovations: an improved backbone network structure enhances feature extraction, an anchorless detection mechanism simplifies model complexity, and a dynamic label assignment strategy improves scenario adaptability. These features enable it to demonstrate performance beyond its predecessor model in flame detection tasks. Looking towards the future, there is still room for optimization of the YOLOv8-based fire detection system. Multimodal data fusion technology can combine visible and infrared images to build a more comprehensive fire characterization system. At the same time, it is necessary to improve the generalization ability of the model in extreme scenarios such as smoke obscuration and develop adaptive optimization algorithms to adapt to different environments. From the application point of view, large-scale deployment of YOLOv8 requires the establishment of a complete solution, including model compression technology, efficient data preprocessing process, and intelligent warning and decision-making system. Its modular design provides a good foundation for function expansion. With the popularization of edge computing devices, technology is expected to achieve large-scale applications in industrial safety, forest fire prevention and other fields, making an important contribution to improving public safety. The technology iteration based on deep learning will continue to promote the development and progress in the field of fire detection.

References

[1] Muhammad, K., Ahmad, J., Lv, Z., Bellavista, P., Yang, P., & Baik, S. W. (2018). Efficient deep CNN-based fire detection and localization in video surveillance applications. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 49(7), 1419-1434.

[2] Muhammad, K., Ahmad, J., Mehmood, I., Rho, S., & Baik, S. W. (2018). Convolutional neural networks based fire detection in surveillance videos. Ieee Access, 6, 18174-18183.

[3] Barmpoutis, P., Dimitropoulos, K., Kaza, K., & Grammalidis, N. (2019, May). Fire detection from images using faster R-CNN and multidimensional texture analysis. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8301-8305). IEEE.

[4] Saeed, F., Paul, A., Karthigaikumar, P., & Nayyar, A. (2020). Convolutional neural network based early fire detection. Multimedia Tools and Applications, 79(13), 9083-9099.

[5] Sharma, A., Kumar, R., Kansal, I., Popli, R., Khullar, V., Verma, J., & Kumar, S. (2024). Fire detection in urban areas using multimodal data and federated learning. Fire, 7(4), 104.

[6] Bhamra, J. K., Anantha Ramaprasad, S., Baldota, S., Luna, S., Zen, E., Ramachandra, R., ... & Nguyen, M. H. (2023). Multimodal wildland fire smoke detection. Remote Sensing, 15(11), 2790.

[7] Toulouse, T., Rossi, L., Akhloufi, M. A., Pieri, A., & Maldague, X. (2018). A multimodal 3D framework for fire characteristics estimation. Measurement Science and Technology, 29(2), 025404.

[8] Chaoxia, C., Shang, W., Zhang, F., & Cong, S. (2022). Weakly aligned multimodal flame detection for fire-fighting robots. IEEE Transactions on Industrial Informatics, 19(3), 2866-2875.

[9] Talaat, F. M., & ZainEldin, H. (2023). An improved fire detection approach based on YOLO-v8 for smart cities. Neural Computing and Applications, 35(28), 20939-20954.

[10] Goyal, S., Shagill, M., Kaur, A., Vohra, H., & Singh, A. (2020). A yolo based technique for early forest fire detection. Int. J. Innov. Technol. Explor. Eng, 9, 1357-1362.

[11] Wu, H., Hu, Y., Wang, W., Mei, X., & Xian, J. (2022). Ship fire detection based on an improved YOLO algorithm with a lightweight convolutional neural network model. Sensors, 22(19), 7420.

[12] Cao, L., Shen, Z., & Xu, S. (2024). Efficient forest fire detection based on an improved YOLO model. Visual Intelligence, 2(1), 20.

[13] Yun, B., Zheng, Y., Lin, Z., & Li, T. (2024). FFYOLO: A lightweight forest fire detection model based on YOLOv8. Fire, 7(3), 93.