

Article

A Lightweight and Sustainable UAV-Based Forest Fire Detection Algorithm Based on an Improved YOLO11 Model

Shuangbao Ma ^{1,2} , Yongji Hui ² , Yapeng Zhang ^{1,2} and Yurong Wu ^{2,*}¹ Hubei Key Laboratory of Digital Textile Equipment, Wuhan Textile University, Wuhan 430073, China; 2006118@wtu.edu.cn (S.M.); ypzhang@wtu.edu.cn (Y.Z.)² School of Mechanical Engineering and Automation, Wuhan Textile University, Wuhan 430073, China; 2315053015@wtu.edu.cn

* Correspondence: 2003025@wtu.edu.cn

Abstract

Unmanned aerial vehicle (UAV) forest fire detection is vital for forest safety. However, early-stage UAV fire scenarios often involve small targets, weak smoke signals, and strict onboard resource constraints, which pose significant challenges to existing detectors. To improve the speed and accuracy of UAV forest fire detection, this paper proposes a lightweight fire detection algorithm, AHE-YOLO, specifically designed for UAVs. The proposed method adopts a coordinated lightweight design to improve feature preservation and cross-scale representation under limited computational budgets. Specifically, the Adaptive Downsampling (ADown) module preserves shallow fire-related cues during spatial reduction, improving sensitivity to small flame and smoke targets. The high-level screening-feature fusion pyramid network (HS-FPN) introduces cross-scale attention to promote more discriminative multi-level feature interaction while reducing redundant computation. Furthermore, the Efficient Mobile Inverted Bottleneck Convolution (EMBC) module is employed to improve receptive-field efficiency and feature selectivity under lightweight constraints, further enhancing detection accuracy and inference speed. Finally, the performance of AHE-YOLO is comprehensively evaluated through ablation and comparative experiments on the same dataset. The final experimental results show that YOLO-AHE achieves a mean average precision (mAP) of 94.8% while reducing model parameters by 39.7%, decreasing FLOPs by 27.0%, and shrinking the model size by 36.4%. In addition, its inference speed improves by 16.5%. Beyond detection performance, the proposed framework supports sustainable forest monitoring by enabling early fire warning with reduced computational and energy demands, showing strong potential for real-time deployment on resource-constrained UAV and edge platforms.

Keywords: forest fire detection; UAV; lightweight deep learning; sustainable environmental monitoring; energy-efficient vision; edge deployment



Academic Editor: Xuelong Li

Received: 21 January 2026

Revised: 25 February 2026

Accepted: 28 February 2026

Published: 3 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Early detection of forest fires is critical for reducing ecological damage and economic loss. Common early detection methods of forest fire mainly include manual observation [1], remote sensing [2] and fixed camera monitoring [3]. Due to the weak fire source and very little smoke in the early stage of forest fires, these methods usually have a long response time, limited coverage and high false alarm rate [4]. UAVs can be used for forest fire detection, which can greatly shorten the disaster response time and enlarge the detection coverage,

and the use of deep learning theory can improve the accuracy of forest fire detection [5]. However, due to UAV hardware limitations, forest fire detection models must achieve both high accuracy and a lightweight design to enable deployment on UAV platforms [6]. In particular, early-stage UAV fire scenarios often involve tiny targets, low-contrast smoke, and complex backgrounds, which further increase the difficulty of reliable detection under onboard resource constraints. Therefore, research on lightweight UAV forest fire detection algorithms is especially important. From a sustainability perspective, early forest fire detection plays a critical role in protecting ecological systems, reducing carbon emissions caused by large-scale wildfires, and minimizing long-term socio-economic losses. Efficient UAV-based monitoring systems can support sustainable forest management by enabling rapid response while reducing operational cost and energy consumption. Therefore, the development of lightweight and resource-efficient fire detection algorithms is not only a technical requirement but also an important component of sustainable environmental monitoring strategies.

Recently, many researchers have extensively investigated intelligent algorithms for forest fire detection [7]. For instance, Harkat et al. proposed an early fire detection approach based on wavelet features, using multiresolution wavelet analysis to describe flame characteristics [8]. Strässle et al. [9] devised a deep learning segmentation algorithm that effectively and consistently delineates flame fronts from combustion photos with low signal-to-noise ratios (SNRs). Seydi et al. [10] introduced the Fire-Net framework, which uses deep convolutional networks to extract spatial information from time-series images, enabling accurate identification of active fire locations. However, these approaches are highly dependent on the background and prone to false positives in challenging environments.

Researchers have made numerous improvements to YOLO models with the intent to enhance forest fire detection [11–13]. Talaat and ZainEldin [14] proposed a YOLOv8-based fire detection model for smart cities, incorporating an attention mechanism and multi-scale feature fusion, which achieved high detection accuracy; however, its computational demands limit real-time deployment. Li et al. [15] proposed LEF-YOLO, a lightweight YOLO-based method for intelligent detection of extreme wildfires, which balances detection accuracy with computational efficiency; however, its performance under highly complex forest conditions still requires further validation. Chen et al. [16] developed the LMDFS model using YOLOv7 to improve smoke feature extraction through coordinate attention and content-aware up-sampling, achieving higher accuracy compared to the baseline. However, accurately detecting small and unpredictable fire patterns remained challenging. Nie et al. [17] proposed a framework that integrates the three-sphere model with YOLOv5 to address class imbalance in remote sensing images. This approach combines spatial, texture, and color information to improve the capability of this model to recognize fire features. Experimental results demonstrated superior detection accuracy under data imbalance conditions compared to the baseline. Despite these advances, existing YOLO-based methods still face difficulties in early-stage UAV scenarios, where targets are extremely small and onboard resources are limited.

The diminished spatial resolution restricts the identification of intricate details in the initial stages of a fire, while traditional satellite monitoring techniques typically depend on a singular thermal infrared band in remote sensing, rendering them susceptible to cloud obstruction [18]. Researchers have recently addressed these limitations by integrating visible, near-infrared, and thermal infrared data fusion approaches to enhance overall system robustness and detection accuracy [19]. Jiang et al. [20] introduced a processing framework utilizing the YOLO model that concurrently examined UAV thermal infrared and visible light data to significantly enhance the detection of low-signal fire sources. Oliver et al. [21] introduced a machine learning-based segmentation technique to precisely identify water

bodies in thermal infrared data, which were frequently misclassified as fire sources due to their significant radiative contrast at night. Utilizing a random forest classifier that incorporates spatial texture and contextual information significantly diminished false alarms and enhanced the precision of tactical wildfire mapping in intricate thermal settings.

Developing lightweight models is crucial for real-time fire detection in resource-limited systems like UAVs [22]. Geng et al. [23] developed YOLOFM, a lightweight YOLOv5n-based fire and smoke detection model optimized for resource-limited environments by reducing redundancy and improving feature fusion.

Moreover, the aerial perspective provided by the UAV reveals that flame targets are typically diminutive and exhibit chaotic patterns, necessitating enhanced small target recognition capabilities [24]. Aral et al. [25] proposed a lightweight and attention-based CNN architecture for wildfire detection using UAV vision data, which integrates attention mechanisms with a compact backbone to reduce model size and adapt to limited onboard hardware. Wang et al. [26] developed a weakly supervised segmentation method for UAV-captured forest fire photos by employing foreground-aware pooling and context-aware loss, achieving precise fire region extraction with minimal manual annotation.

Although considerable progress has been made, the joint requirements of small-target sensitivity, real-time performance, and UAV onboard efficiency are still not fully satisfied by existing lightweight detectors. For UAV-based fire monitoring, the detector must achieve real-time performance under limited onboard resources. The YOLO family is selected as the primary baseline because it offers a favorable speed–accuracy trade-off and a streamlined one-stage detection pipeline, which has been widely validated in real-time embedded vision tasks. This practical advantage makes YOLO-based architectures well suited for deployment-oriented UAV fire detection. Therefore, this paper introduces a lightweight fire detection algorithm, AHE-YOLO, specifically designed for accurate identification of small flames and smoke in UAV imagery. Unlike conventional lightweight modifications that rely on isolated module replacement, the proposed method adopts a coordinated lightweight design to improve feature preservation and cross-scale representation under strict resource constraints. This design is specifically tailored to the characteristics of early-stage UAV fire imagery and is not a simple architectural recombination. As a result, AHE-YOLO achieves an improved balance between real-time performance and detection accuracy while maintaining a compact model structure. This work primarily presents the following contributions:

- (1) A lightweight UAV forest fire detection algorithm, AHE-YOLO, is proposed, and its performance is evaluated.
- (2) Ablation experiments and comparative experiments are conducted to demonstrate the excellent performance of the proposed algorithm, compared with RT-DETR, YOLOv5, YOLOv8 and YOLOv10, in order to emphasize the characteristics of the improved algorithm in forest fire detection, such as its precision and lightweight design.
- (3) Through the comparative analysis of the actual measurement data of UAV forest fire detection algorithm, it can be seen that AHE-YOLO has powerful real-time ability and edge device adaptability, and can detect small flames and smoke objects with high precision while maintaining the compactness of the model.

Compared with existing YOLO-based lightweight models, AHE-YOLO incorporates a multi-path downsampling strategy, a cross-scale fusion neck, and an efficient attention-enhanced bottleneck, achieving both high accuracy and real-time performance for UAV-based fire detection.

The rest of this article is organized as follows. Section 2 briefly introduces the related principles and the improved algorithm AHE-YOLO. In Section 3, we elaborate on the performance of this improved algorithm, the ablation experiment, and the compar-

ison experiment. Section 4 reports the comparative analysis of the actual measurement data of UAV forest fire detection algorithm. Section 5 concludes this paper and presents future work.

2. Lightweight Forest Fire Detection Algorithm AHE-YOLO

In UAV forest fire detection, the detection algorithm must balance high accuracy with a lightweight model design [27]. Based on the characteristics of fire detection, this paper proposes an improved lightweight algorithm, AHE-YOLO, built on the YOLO11 network, whose overall structure is illustrated in Figure 1.

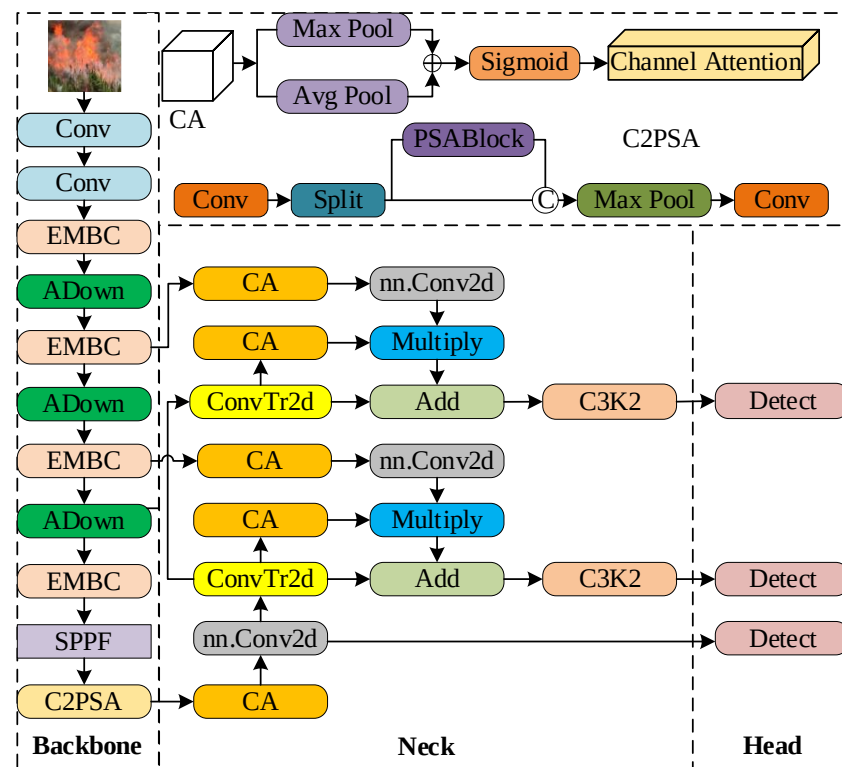


Figure 1. The structural diagram of YOLO-AHE.

In this work, YOLO11 denotes the real-time detector implemented within the Ultralytics framework, rather than a newly proposed detection architecture, and serves as the baseline architecture for all ablation and comparative experiments. We adopt this configuration because it provides a lightweight and stable starting point under a unified training and inference pipeline, thereby enabling a fair and controlled evaluation of the proposed ADown, HS-FPN, and EMBC modules.

To better address early-stage UAV fire scenarios, the proposed framework introduces several targeted enhancements to the baseline architecture. Rather than relying on isolated lightweight operations, AHE-YOLO adopts a coordinated design strategy to improve small-target perception under strict computational constraints. Specifically, ADown is designed to preserve shallow fire cues during downsampling, HS-FPN strengthens cross-scale feature interaction for small flame and smoke targets, and EMBC enhances feature representation through efficient receptive-field expansion. Among these components, EMBC provides the primary modeling enhancement by improving receptive-field efficiency under lightweight constraints, while ADown and HS-FPN mainly serve task-oriented complementary roles. The overall processing workflow of the proposed method is further described in the following subsection.

2.1. Methodology Overview

Figure 1 illustrates the overall framework of AHE-YOLO. From an end-to-end pipeline perspective, given an input UAV image, the backbone first applies ADown to reduce spatial resolution while preserving shallow fire-related cues. The neck then adopts HS-FPN to perform cross-scale feature selection and fusion, which enhances the representation of small flame and smoke targets. Subsequently, EMBC blocks strengthen feature extraction by providing an efficient receptive-field expansion with low computational overhead. Finally, the detection head outputs the bounding boxes and class probabilities for flame and smoke objects.

2.2. Adaptive Downsampling Module

The ADown module utilizes a hybrid downsampling method to reduce the spatial dimensions of feature maps while preserving essential information through a dual-path feature fusion approach. This design optimizes computational efficiency and detection accuracy [28].

The efficient operation of the ADown module does not significantly affect performance. Minimizing the number of parameters diminishes model complexity and improves operational efficiency, especially in resource-limited settings. The meticulous design of the module retains the largest quantity of image information while reducing the spatial resolution of feature maps, thereby enhancing target identification accuracy. Moreover, ADown possesses learnable characteristics that enable it to adapt dynamically to diverse data configurations and enhance fire detection efficiency [29]. Figure 2 illustrates the structure of the ADown downsampling module.

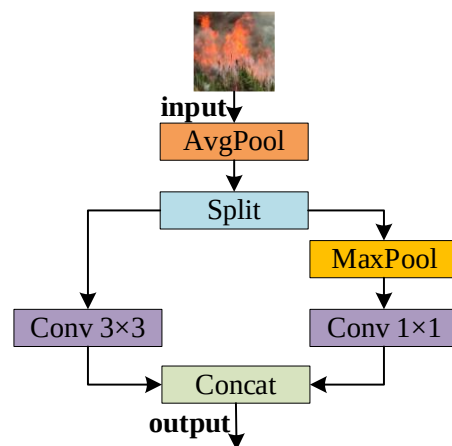


Figure 2. The structural diagram of the ADown downsampling module.

The primary objective of the ADown module is to achieve efficient feature downsampling through structural optimization. The process commences with the implementation of average pooling (kernel size 2, stride 1) on the input feature map, subsequently partitioning the outcome along the channel dimension into two branches. The initial branch undergoes max pooling utilizing a 3×3 kernel size and a stride of 2, followed by a 1×1 convolution. A 3×3 convolution with a stride of 2 processes the second branch. The final downsampled feature map is generated by consolidating the findings from the two branches along the channel dimension. This architecture improves the capacity of the model to detect small objects by effectively reducing spatial resolution while maintaining critical semantic information. This dual-path design helps preserve shallow fire-related cues during downsampling while suppressing redundant background responses, thereby improving feature stability for small and low-contrast targets.

The ADown module reduces both the number of parameters and computational burden while maintaining or enhancing detection accuracy. In UAV forest fire detection, its exceptional efficiency and lightweight nature render it ideal. Furthermore, its adaptable design facilitates modifications to various data configurations, thereby improving overall model performance.

2.3. High-Level Screening-Feature Fusion Pyramid Network

This improved algorithm uses the HS-FPN as the neck component of the model, facilitating fast multi-scale feature fusion and preliminary feature selection. The HS-FPN begins by employing a Channel Attention (CA) module on the input feature maps, enhancing the ability of model to focus on important channels [30].

Figure 3 shows that the CA module performs global average pooling and global max pooling on the input to generate two sets of feature descriptors, which are then fused. The Sigmoid activation function computes the channel-wise attention weights. This method enables the module to accentuate fire-related attributes in each channel while reducing information loss. This attention-guided mechanism promotes more discriminative cross-scale feature allocation, allowing the network to emphasize small and weak fire responses while suppressing background interference. The channel weights are summed element-wise with the original feature maps at each scale, producing attention-weighted outputs. Prior to feature fusion, a 1×1 convolution is employed to equally reduce the channel count of each scale feature map to 256, hence ensuring consistency in the channel dimension across diverse scale feature maps.

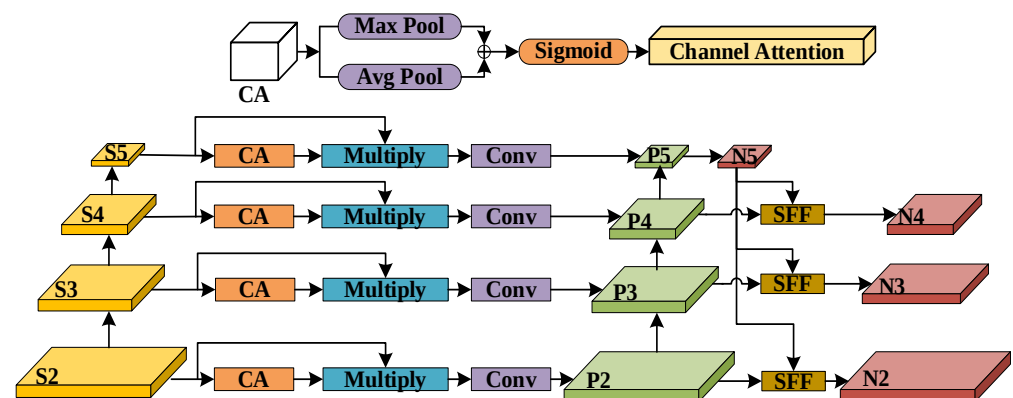


Figure 3. The structural diagram of the HS-FPN.

Feature maps are generated from high-level features and low-level features by executing a transpose-convolution operation, utilizing a kernel size of 3×3 and a step size of 2. To ensure consistent dimensions between high-level and low-level features, a bilinear interpolation up-sampling (or downsampling) operation is applied to the high-level features, resulting in dimensionally matched features. The low-level feature map, filtered by the CA module, is fused with the aligned high-level feature map to produce the output feature. The computational procedure of the Selective Feature Fusion (SFF) module is illustrated in Equations (1) and (2).

$$f_{att} = BL(T - Conv(f_{high})) \quad (1)$$

$$f_{out} = f_{low} \times CA(f_{att}) + f_{att} \quad (2)$$

The HS-FPN module selectively integrates forest fire-relevant feature maps using the Channel Attention (CA) mechanism and the Dimension Combining (DM) method. The CA module extracts key information from each channel through pooling and weight computation. Subsequently, the SFF module integrates the previously extracted fire-related

components, as illustrated in Figure 4. This strategy improves the detection of small-scale fire targets, such as localized flames, while considering other fire-related features, enhancing the detection capability of the model and adaptability in complex fire scenarios. Moreover, the streamlined HS-FPN architecture reduces computational redundancy and improves efficiency, providing key design insights for the improvements proposed in this study.

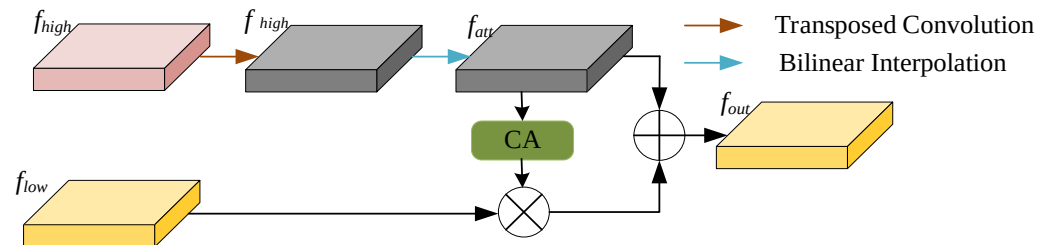


Figure 4. Structural diagram of the SSF module.

2.4. Efficient Mobile Inverted Bottleneck Convolution

This improved algorithm utilizes the EMBC module to replace the C3K2 module in the YOLO11 backbone in order to improve the detection efficacy of UAV forest fire detection. The EMBC module enhances feature extraction and representation, significantly improving detection accuracy and speed.

Figure 5 shows that the MBConv module employs the inverted residual architecture derived from MobileNetV3 [31]. Enhancements have been made by the incorporation of an attention mechanism and a Depthwise Separable Convolution (DSC) methodology, among other modifications. The former enhances feature selectivity, whereas the latter significantly diminishes the computational load of the model and the parameter quantity of the model, thereby augmenting the generalizability of the model and decreasing the risk of overfitting. This renders the model suitable for resource-constrained platforms, such as UAVs [32].

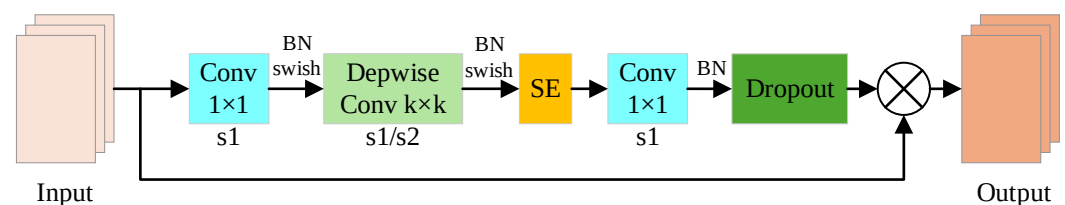


Figure 5. The structural diagram of MBConv.

Figure 5 illustrates the intricate architecture of MBConv. The initial step involves up-sampling by 1×1 convolution, succeeded by BatchNorm normalization with Swish activation, and subsequently, spatial features are restored using $k \times k$ depth-separable convolution. The Squeeze-and-Excitation (SE) module enhances channel weight perception, dimensionality reduction is achieved using 1×1 convolution, and a Dropout layer is incorporated to prevent overfitting [33].

The EfficientNet framework, created by the Google AI team, presents a unified composite scaling method that aligns the depth and width of the network with the input resolution to improve performance. This work introduces the EMBC module to improve the original MBConv with a more efficient channel attention technique called EffectiveSE. This module first expands the channel dimension to four times its original size using 1×1 convolution, then applies the mean method to compute the channel average without diminishing dimensionality, thereby avoiding the information loss typically associated with the standard SE module during dimensionality reduction. The channel weights produced by the hard-sigmoid activation function enhance the adaptive capability among feature

channels. This design improves receptive-field efficiency under lightweight constraints, enabling more effective capture of small and spatially sparse fire patterns while maintaining low computational overhead. The EMBC module replaces the ReLU and Hardsigmoid activation functions with the less resource-intensive Swish function in the original MBConv to decrease computational complexity. Figure 6 illustrates the structure of the EMBC module.

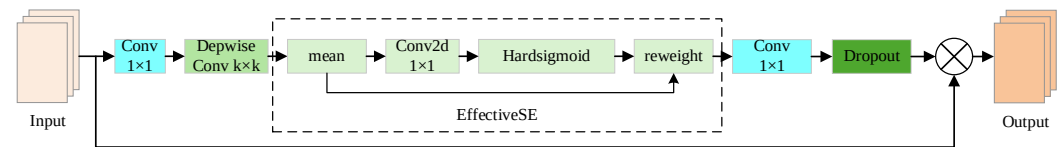


Figure 6. The structural diagram of the EMBC module.

The EMBC module enhances the capability of the model to extract forest flame properties, hence facilitating the development of a lightweight, high-performance target detection system and considerably improving the operating efficiency of the network on edge devices.

Based on the above architectural design, comprehensive experiments are conducted to systematically evaluate the effectiveness and efficiency of the proposed AHE-YOLO, as presented in Section 3. Overall, the proposed modules enhance feature representation while preserving strict lightweight characteristics.

3. Experiments and Analysis

3.1. Dataset

The dataset used in this study was collected from screenshots of open-source Roboflow resources, social media, news reports, and keyframes extracted from public surveillance videos, and all images were manually annotated using LabelImg v1.8.1. The final dataset contains 5118 annotated images (4096 for training, 511 for validation, and 511 for testing), all of which include at least one fire or smoke instance. The collected data cover a variety of complex natural scenes, including dense forests, wildfire environments, and nighttime fire conditions. In the experiments, the dataset was divided into training, validation, and test sets with a ratio of 8:1:1 for model training, parameter tuning, and performance evaluation. This split helps ensure the reliability of the experimental results and supports the assessment of model generalization under different scenarios. Some representative samples are shown in Figure 7. To better reflect real UAV monitoring conditions, the dataset covers a wide range of fire scenarios, including smoke-dominant cases, small or low-contrast fire targets, and challenging environments with complex illumination or atmospheric interference. This diversity enables a more comprehensive evaluation of detection sensitivity.

It should be noted that the current dataset mainly consists of fire- and smoke-related positive samples, and no pure negative images without fire or smoke are included. This setting is consistent with many early fire detection studies that focus on improving sensitivity to fire-related targets. Meanwhile, the robustness of the proposed method is evaluated through diverse background conditions and cross-scenario testing. In future work, we will further expand the dataset by incorporating additional negative samples to strengthen false-alarm evaluation under broader real-world conditions.



Figure 7. Representative samples of the constructed forest fire dataset, covering diverse UAV monitoring conditions, including smoke-dominant scenes, small and distant fire targets, nighttime fires, and complex illumination environments (e.g., dusk and backlighting).

To provide qualitative insight into the detection performance, Figure 8 presents visual comparisons between manual ground-truth annotations and the predictions of AHE-YOLO. As shown in the figure, the proposed method can accurately localize most fire and smoke regions under diverse UAV forest scenarios, including dense smoke, large-scale flames, and complex backgrounds.

In most cases, the predicted bounding boxes closely match the ground-truth annotations, indicating reliable spatial localization capability. However, slight localization deviations can still be observed in some challenging scenes, particularly when smoke boundaries are diffuse or when flames are partially occluded by vegetation. These observations suggest that while AHE-YOLO demonstrates strong robustness, further improvements could be achieved through more fine-grained feature modeling and the inclusion of harder training samples in future work. Importantly, we acknowledge that the absence of pure negative scenes may limit comprehensive false-alarm assessment, and this aspect will be further improved in future dataset extensions.

In addition, the try123-v4 dataset is employed for forest fire detection experiments. This dataset is sourced from the Roboflow platform and contains 2967 images covering multiple stages of fire development, ranging from early smoke diffusion to fully developed flames. It is used in this study to evaluate the generalization performance of the proposed model and to assess its adaptability under different fire stages and complex environmental conditions.

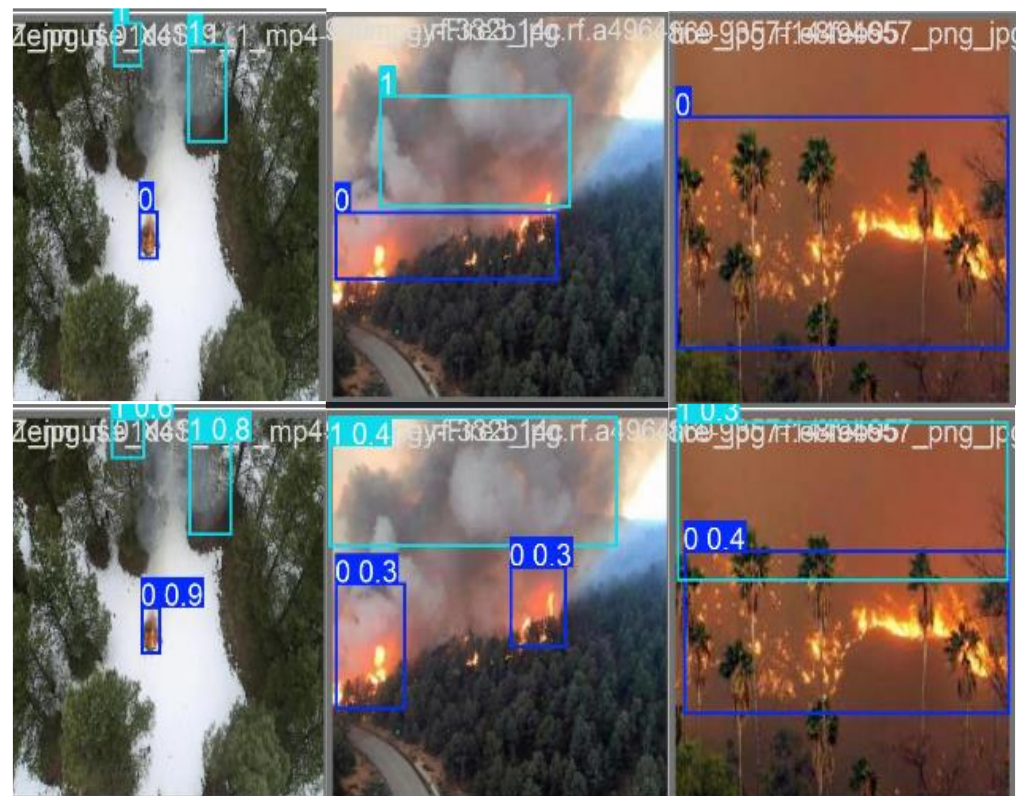


Figure 8. Visual comparison between ground-truth annotations (**top row**) and AHE-YOLO detection results (**bottom row**) across representative UAV forest fire scenarios. Note: Dark blue boxes denote detected flames and light blue boxes denote detected smoke. The label above each bounding box is formatted as “class confidence”, where “0” denotes flame, “1” denotes smoke, and the decimal value (e.g., 0.9) represents the detection confidence.

3.2. Training Equipment and Parameter Setting

All experiments in this study are conducted on a high-performance Windows workstation (Dell Precision 7920 Tower). The system is equipped with two Intel Xeon Gold 6128 CPUs (3.40 GHz, 12 cores in total), 112 GB DDR4 RAM (2666 MHz), and an NVIDIA Quadro P6000 GPU with 24 GB memory, providing reliable computational support for deep learning training and forest fire detection tasks.

The implementation is based on Python 3.11.5 and PyTorch 2.1.1, with CUDA 11.8 enabled for GPU acceleration. All input images are resized to 640×640 .

Unless otherwise specified, all models are trained under identical experimental settings for fair comparison. The training schedule consists of 200 epochs with a batch size of 16, using Stochastic Gradient Descent (SGD) as the optimizer. An early stopping strategy is applied to prevent overfitting, terminating training if no performance improvement is observed for 50 consecutive epochs. The detailed hyperparameter configuration follows Table 1 and is derived from commonly adopted YOLO settings, which are further verified through preliminary experiments to ensure stable convergence.

During inference evaluation, experiments are performed on the same NVIDIA Quadro P6000 GPU with the input resolution fixed at 640×640 and the batch size set to 1. The reported FPS values are averaged over multiple forward passes after model warm-up. Fixed random seeds are used throughout all experiments to reduce performance variance. In addition, repeated pilot runs show consistent convergence trends across models, indicating that the observed performance gains are stable rather than caused by training randomness. These results suggest the deployment potential of the proposed model on

resource-constrained platforms, although further validation on embedded or UAV-class hardware will be conducted in future work.

Table 1. Training parameter settings.

Parameter	Setting	Parameter	Setting
Epochs	200	patience	50
imgsz	640	momentum	0.937
batch	32	weight_decay	0.0005
workers	4	lr0	0.01
optimizer	SGD	lrf	0.01

3.3. Model Evaluation Metrics

To comprehensively evaluate the performance of the YOLO-AHE model in forest fire detection, this study considers two dimensions of evaluation: detection accuracy and model complexity. For detection accuracy, the primary metrics are mean average precision (mAP) and recall (R). Specifically, mAP is calculated as the area under the Precision–Recall (PR) curve for each object class, reflecting the model’s overall detection accuracy across different recall levels. Recall measures the model’s ability to identify all true-positive instances, indicating the extent of missed detections.

For model complexity, this study evaluates the model’s lightweight characteristics using four metrics: floating point operations (FLOPs), number of parameters, model size, and inference speed measured in frames per second (FPS).

3.4. Performance Analysis of the AHE-YOLO Model

Figure 9 illustrates the normalized confusion matrix of the model on the dataset, showing trends in misclassification and classification accuracy. The model demonstrates strong performance in smoke detection, achieving an accuracy of 92%, indicating robust detection capability. Flame detection accuracy is 83%, with a 17% false detection rate, mainly due to misclassification as background.

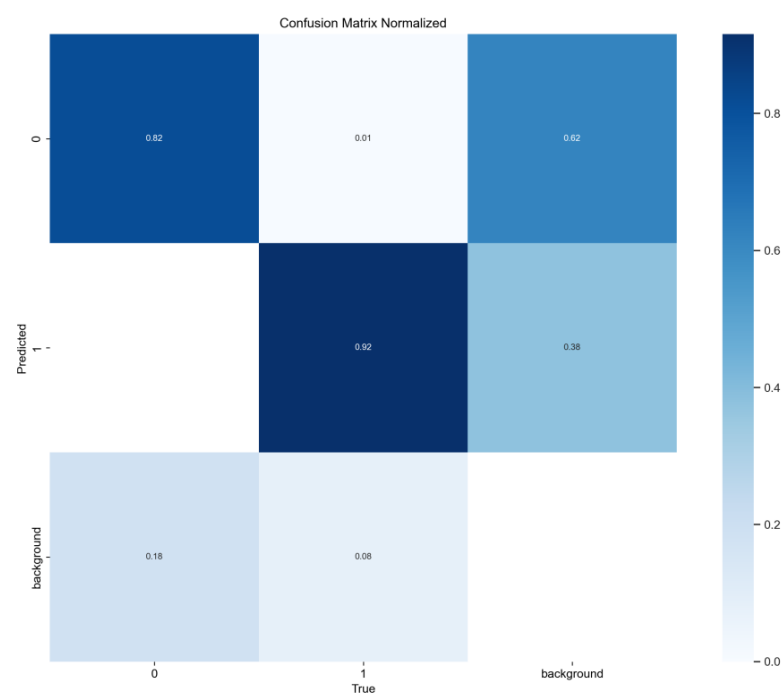


Figure 9. Normalized confusion matrix. Note: In the normalized confusion matrix, class “0” denotes flame and class “1” denotes smoke. The diagonal elements represent correct classifications, while the off-diagonal elements indicate misclassifications between categories.

Figure 10 illustrates the training curve of the AHE-YOLO model for the forest fire recognition issue presented. Following about the 200th epoch, the model enters the stabilization phase; the entire convergence process is both rapid and seamless, demonstrating robust training efficiency and exceptional feature extraction capability. In complex natural environments, AHE-YOLO effectively detects flame-related regions while reducing the risk of overfitting.

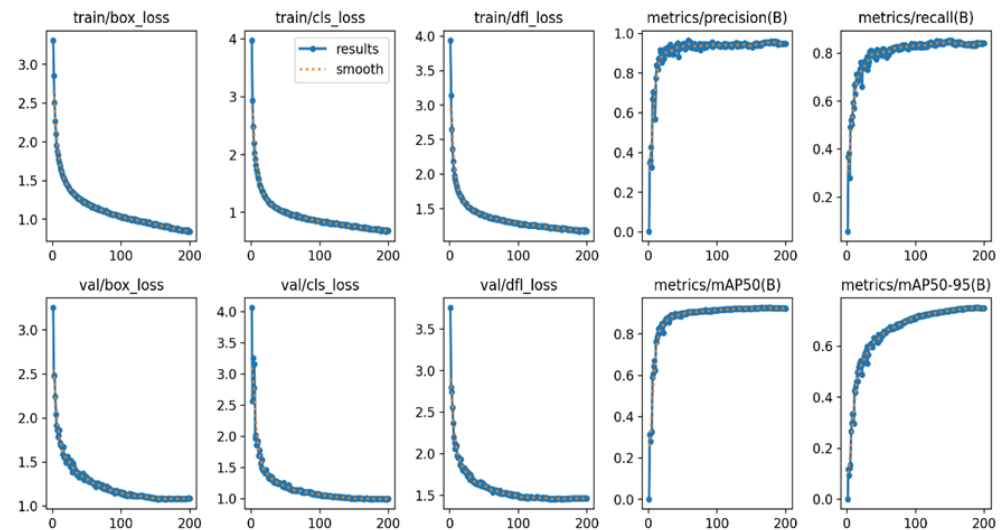


Figure 10. The convergence graph of model training.

3.5. Ablation Experiment

Ablation experiments are conducted utilizing a custom fire dataset to meticulously examine and verify the influence of each model component on overall performance. All trials were conducted under consistent training circumstances and parameters. The conclusive outcomes of the ablation experiment are presented in Table 2.

Table 2. Comparison of alation experiments.

Network	ADown	HS-FPN	EMBC	mAP	Recall	Parameters	FLOPS	Model Size	FPS
YOLO11	-	-	-	92.1%	83.4%	2,582,542	6.3 G	5.5 MB	285.7 frame·s ⁻¹
Module 1	✓	-	-	92.6%	83.9%	2,103,310	5.3 G	4.5 MB	285.7 frame·s ⁻¹
Module 2	-	✓	-	92.8%	84.4%	1,837,558	5.9 G	4.0 MB	294.1 frame·s ⁻¹
Module 3	-	-	✓	93.6%	83.6%	2,875,494	6.1 G	6.1 MB	288.7 frame·s ⁻¹
Module 4	✓	✓	-	93.1%	83.8%	1,491,446	5.0 G	3.3 MB	294.7 frame·s ⁻¹
Module 5	-	✓	✓	93.5%	84.2%	1,929,502	5.8 G	4.2 MB	303.0 frame·s ⁻¹
Module 6	✓	-	✓	94.1%	84.5%	2,195,254	5.2 G	4.8 MB	312.5 frame·s ⁻¹
Ours	✓	✓	✓	94.8%	85.1%	1,557,910	4.6 G	3.5 MB	333.3 frame·s ⁻¹

Note: The symbol “✓” denotes that the improvement module has been incorporated into the model, while the symbol “-” indicates that this improvement has not been implemented.

Table 2 demonstrates the systematic evaluation of the ADown, HS-FPN, and EMBC modules in terms of their effects on model performance and computational efficiency. The experimental results indicate that these modules enhance parameter efficiency and reduce computational burden while simultaneously improving detection accuracy.

The ADown downsampling structure, as shown in Model 1, effectively preserves semantic features and minimizes redundant computation. With parameters reduced to 2.10 million and FLOPs decreased to 5.3 billion, the mAP rises to 92.6%, and the model size is reduced to 4.5 MB. Model 2 improves multi-scale feature detection by using the HS-FPN module, which combines an attention mechanism with a feature fusion network.

The parameters are decreased to 1.84 million, yielding a model size of only 4.0 megabytes. Concurrently, mAP increases to 92.8%, and recall improves to 84.4%, signifying considerable efficacy. Model 3 utilizes the EMBC module to enhance feature representation depth and receptive field, achieving a 93.6% mAP and 83.6% recall. Consequently, this leads to an augmentation in parameters (2.88 M), FLOPs (6.1 G), and model size (6.1 MB).

Model 4 integrates HS-FPN with ADown. The mAP is marginally improved to 93.1%, with the number of parameters reduced to 1.49 million and FLOPs decreased to 5.0 billion. The model size was constrained to 3.3 MB, highlighting substantial lightweight advantages. Model 5 combines HS-FPN and EMBC to enhance semantic understanding and enable deeper feature extraction. The model achieves a mAP of 93.5% and a recall of 84.2%, contains 1.93 million parameters, maintains modest computational resource usage, and has an overall size of 4.2 MB. Model 6 demonstrates the combination of ADown and EMBC, achieving a mAP of 94.1% and a recall of 84.5% by integrating information preservation with advanced feature extraction techniques. This configuration ranks among the highest-performing models, comprising 2.20 million parameters, 5.2 G FLOPs, and a model size of 4.8 MB.

The final YOLO-AHE model (ADown + HS-FPN + EMBC) achieves an optimal balance between accuracy and efficiency: a mAP of 94.8% and a recall of 85.1%; meanwhile, the number of parameters is reduced to 1.56 M, FLOPs to 4.6 G, and the model size to only 3.5 MB, with an inference speed of 333.3 frames·s^{−1}. These results indicate that the model achieves excellent deployment efficiency while maintaining high detection accuracy, demonstrating significant practical application potential.

Although the baseline already achieves real-time performance on high-end GPUs, practical UAV platforms are typically constrained by onboard computing capability, memory capacity, and power consumption. Therefore, reducing model size and computational overhead remains essential for reliable edge deployment.

3.6. Comparative Experiments

This paper systematically evaluates the generalization capability and robustness of the AHE-YOLO model in target detection by conducting comparative experiments with other leading detection methods and architectures. The two-stage RT-DETR is chosen as a representative model among Transformer-based architectures for comparison purposes. Moreover, numerous widely used one-stage detection methods, such as YOLOv5, YOLOv8, and YOLOv10, are demonstrated to evaluate the performance discrepancies of AHE-YOLO across different lightweight methodologies with conventional architectures. All models are systematically trained and assessed using the self-constructed forest fire image dataset; results are presented in Table 3.

Table 3. Comparative experiment of algorithms.

Network	mAP	Recall	Parameters	FLOPS	Model Size	FPS
RT-DETR-l	88.9%	80.3%	28,447,370	105.2 G	59.1 MB	51.5 frame·s ^{−1}
YOLOv5n	91.6%	84.1%	2,182,054	5.8 G	4.7 MB	245.7 frame·s ^{−1}
YOLOv5s	91.3%	83.9%	7,814,390	18.7 G	16.0 MB	107.5 frame·s ^{−1}
YOLOv8n	91.9%	84.4%	2,684,758	6.8 G	5.6 MB	253.2 frame·s ^{−1}
YOLOv8s	91.6%	84.6%	9,839,734	23.6 G	19.0 MB	158.6 frame·s ^{−1}
YOLOv10n	91.5%	81.7%	2,265,558	6.5 G	5.8 MB	270.2 frame·s ^{−1}
YOLOv10s	91.1%	83.9%	7,218,774	21.4 G	16.5 MB	153.8 frame·s ^{−1}
YOLO11	92.1%	83.4%	2,582,542	6.3 G	5.5 MB	285.7 frame·s ^{−1}
AHE-YOLO	94.8%	85.1%	1,557,910	4.6 G	3.5 MB	333.3 frame·s ^{−1}

All baseline models are implemented using their official configurations and retrained on our dataset under identical experimental settings to ensure fairness.

As shown in Table 3, YOLO-AHE achieves the best performance across all evaluation metrics, with a mAP of 94.8% and a recall of 85.1%. Meanwhile, its model has only 1.56 M parameters, requires 4.6 GFLOPs of computation, and occupies just 3.5 MB of memory. Compared with commonly used models, YOLO-AHE's mAP is 3.2% and 2.9% higher than YOLOv5n and YOLOv8n, respectively, while its number of parameters is only 71.4% and 58.1% of those models. Moreover, it achieves an FPS of 333.3 frames·s⁻¹, reaching the highest detection speed. These results demonstrate that YOLO-AHE strikes a favorable balance among detection accuracy, computational complexity, model size, and inference speed, highlighting its strong application potential.

This paper standardizes the data in Table 3 to facilitate a straightforward comparison of the overall performance of each model across several critical parameters. These metrics are visualized using radar plots, which include mAP, Recall, Parameters, FLOPs, and Model Size, clearly highlighting the strengths and weaknesses of each model. It is essential to recognize that the normalization procedure is reversed for metrics where smaller values are preferable, such as Parameters, FLOPs, and Model Size. This is implemented to ensure that elevated values consistently indicate superior performance on the radar plots. This method of visualization enables a comprehensive comparison of the detection accuracy and computational efficiency of each model while maintaining a lightweight structure. It may serve as a reference for selecting and implementing models in the future. Figure 11 displays the radar chart.

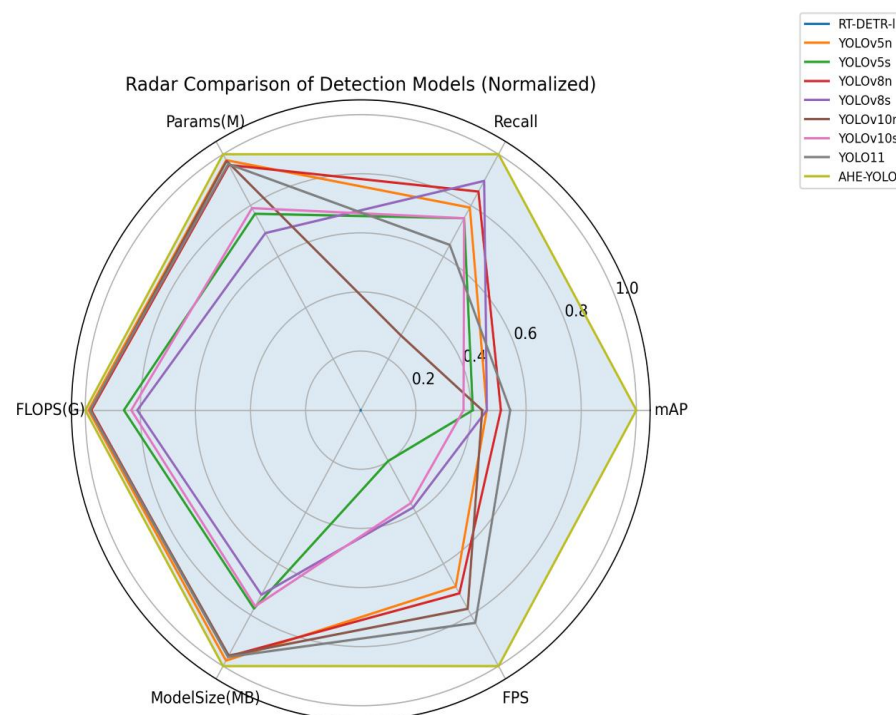


Figure 11. Radar comparison plot of different and improved models.

To verify that the proposed ADown module achieves model lightweighting while maintaining detection accuracy, comparative experiments were conducted using several mainstream convolutional variants, including LAWDS, WaveletPool, ContextGuide Down, and the ADown module introduced in this study. Under the same network architecture and training conditions, the ADown module attains a mAP of 92.6% and a recall of 83.9% while reducing the number of parameters to 2.10 M, FLOPs to 5.3 G, and model size to only

4.5 MB. These results demonstrate a favorable balance between performance and efficiency. The corresponding comparison results are summarized in Table 4.

Table 4. Comparative experiments of different convolution improvement modules.

Network	mAP	Recall	Parameters	FLOPS	Model Size
LAWDS	92.1%	82.9%	2,240,398	6.4 G	4.8 MB
WaveletPool	92.6%	83.7%	2,168,286	5.4 G	4.7 MB
ContextGuide Down	92.8%	83.7%	3,530,128	9.0 G	7.5 MB
ADown	92.6%	83.9%	2,103,310	5.3 G	4.5 MB

Compared with the LAWDS module, ADown achieves higher accuracy. Compared with the WaveletPool module, which has similar performance, ADown maintains the same accuracy while requiring fewer parameters and lower FLOPs. Although the ContextGuide Down module achieves a slightly higher mAP, its parameter count reaches 3.53 M, its FLOPs reach 9.0 G, and its model size increases significantly, making it less suitable for deployment on small devices. In summary, compared with mainstream modules, ADown offers a superior overall balance between lightweight design and detection accuracy, making it more suitable for deployment on edge devices.

The effectiveness of HS-FPN for flame detection was evaluated through comparative experiments using several mainstream feature pyramid structures, with results summarized in Table 5. The mAP scores of the different structures are as follows: BI-FCN 91.9%, CS-FCN 92.1%, G-FPN 92.4%, and the proposed HS-FPN 92.8%. HS-FPN slightly outperforms G-FPN in accuracy, achieving a recall of 84.4% and demonstrating excellent object detection performance. In addition, HS-FPN has only 1.84 M parameters, 5.9 G FLOPs, and a model size of 4.0 MB, with an mAP clearly higher than both G-FPN and CS-FCN. More importantly, HS-FPN combines high detection accuracy with a lightweight design, making it highly suitable for practical deployment in UAV vision and edge computing applications.

Table 5. Comparison experiments of different feature pyramid improvement modules.

Network	mAP	Recall	Parameters	FLOPS	Model Size
BI-FCN	91.9%	83.2%	1,923,018	6.3 G	4.2 MB
CS-FCN	92.1%	82.2%	2,964,382	7.2 G	6.3 MB
G-FPN	92.4%	83.6%	3,659,998	8.2 G	7.8 MB
HS-FPN	92.8%	84.4%	1,837,558	5.9 G	4.0 MB

To further validate the feature extraction capability of EMBC, comparative experiments were conducted with current mainstream C3K2-based improvements, including KAN, DynamicConv, and FMB, with results summarized in Table 6. The experiments show that the EMBC module, while maintaining a lightweight structure, achieves a mAP of 93.6% and a recall of 83.6%, with only 2.88 M parameters, 6.1 G FLOPs, and a model size of 6.1 MB. Overall, EMBC outperforms other mainstream feature extraction methods in terms of both efficiency and accuracy.

Table 6. Comparative experiment of the C3K2 improvement module.

Network	mAP	Recall	Parameters	FLOPS	Model Size
KAN	93.2%	83.7%	3,319,470	6.3 G	7.0 MB
DynamicConv	93.0%	85.2%	3,460,850	6.1 G	7.3 MB
FMB	91.3%	80.6%	2,647,870	6.7 G	5.7 MB
EMBC	93.6%	83.6%	2,875,494	6.1 G	6.1 MB

Although the KAN and DynamicConv modules achieve relatively high mAP scores of 93.2% and 93.0%, respectively, their parameter counts exceed 3.3 M. On the other hand, while the FMB module has a smaller parameter count of only 2.65 M, its detection accuracy is lower, with a mAP of 91.3% and a recall of 80.6%. In contrast, the EMBC module not only improves detection accuracy but also minimizes both parameter count and model size compared with the other modules, demonstrating a clear advantage.

3.7. Cross-Dataset Validation on Try123-v4

To further evaluate the generalization capability of the proposed method and mitigate potential bias introduced by the self-constructed dataset, additional experiments are conducted on the public try123-v4 dataset. For fair comparison, the proposed model is retrained under settings that are consistent with those reported in the comparative studies.

On the try123-v4 dataset, Ref. [34] introduced a bidirectional feature pyramid network (BiFPN) and designed the C2f_MLCA module while replacing the Complete Intersection over Union (CIoU) loss with the Inner-DIoU loss. Ref. [35] enhanced multi-scale extraction of high- and low-frequency smoke features by integrating wavelet convolution and further strengthened attention to smoke regions using the SimAM mechanism.

As shown in Table 7, although Refs. [34,35] achieve slightly higher detection accuracy, the proposed method significantly reduces the number of parameters and computational complexity while maintaining competitive performance. These results indicate that AHE-YOLO provides a more favorable trade-off between accuracy and lightweight efficiency, demonstrating good generalization ability across different datasets.

Table 7. Cross-dataset comparison results on the try123-v4 dataset.

Dataset	Method	mAP	Recall	Parameters	FLOPS	Model Size
try123-v4	Ours	78.9%	70.7%	1,557,910	4.2 G	3.5 MB
	Ref. [34]	79.5%	70.3%	1,991,754	7.1 G	4.3 MB
	Ref. [35]	79.2%	70.1%	2,754,000	6.6 G	-

4. Discussion

4.1. Visual Comparison of Detection Results

Figure 12 illustrates the comparative detection performance of YOLO11 and the enhanced AHE-YOLO model across various typical forest fire scenarios. The first row presents the detection results produced by YOLO11, while the second row shows the corresponding results of AHE-YOLO. As observed, AHE-YOLO exhibits more robust and consistent detection behavior across diverse scenarios. Overall, the proposed design emphasizes efficiency-aware feature learning, which aligns with the practical requirements of real-world UAV fire monitoring systems. The proposed lightweight architecture reduces computational burden and energy demand, which is particularly important for long-duration UAV missions and sustainable monitoring systems operating in remote forest regions.

While AHE-YOLO demonstrates superior accuracy and higher confidence in detecting small, open flames, reflecting increased sensitivity to low-contrast targets, Figure 12a,e indicate that YOLO11 suffers from partial detection failures and exhibits lower confidence in densely forested regions. Figure 12b,f further show that AHE-YOLO generates more comprehensive bounding boxes around peripheral flames and achieves enhanced confidence levels, thereby reducing both false detections and missed detections compared to YOLO11. This improvement is particularly significant in mountainous fire scenarios, where strong light–dark contrasts and smoke occlusion pose substantial challenges for accurate detection.

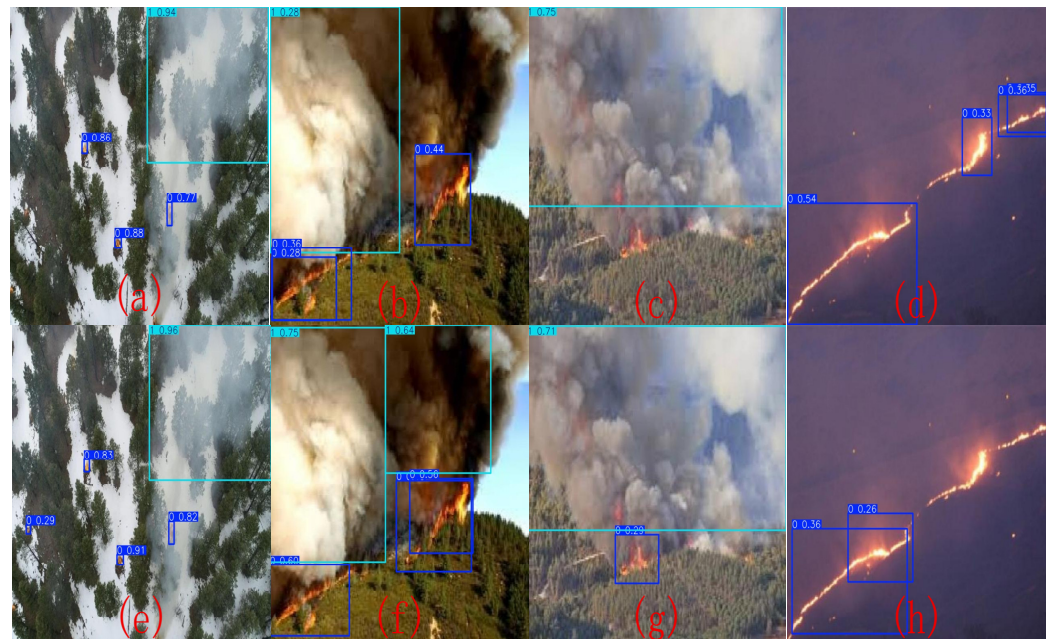


Figure 12. Visual comparisons of detection performance: (a,e) Areas deep within dense forests, (b,f) Mountain wildfires, (c,g) Distant fire spots, (d,h) Open flames at night. Note: Dark blue boxes denote detected flames and light blue boxes denote detected smoke. The label above each bounding box is formatted as “class confidence”, where “0” denotes flame, “1” denotes smoke, and the decimal value (e.g., 0.36) represents the detection confidence.

While YOLO11 fails to detect a tiny open flame at the bottom of the image, AHE-YOLO successfully recognizes it amidst the dense smoke and distant fire regions depicted in Figure 12c,g, thereby demonstrating its perceptual effectiveness in complex backgrounds. Figure 12d,h demonstrate that YOLO11 displays diminished confidence and erroneous detections in nocturnal fire scenarios characterized by low ambient illumination. In contrast, AHE-YOLO demonstrates enhanced adaptability to low-light environments and improved anti-interference capability by effectively distinguishing fire sources from background highlights.

The AHE-YOLO model demonstrates higher detection accuracy and improved robustness compared to the baseline YOLO11 across diverse scenarios, particularly excelling in challenging conditions such as smoke occlusion, complex backgrounds, and nighttime fires.

In comparison with YOLO11, AHE-YOLO exhibits stronger activation responses in flame and smoke regions while effectively suppressing background interference from trees and terrain. In dense forest scenes, the proposed model clearly highlights the flame contours, thereby enhancing localization accuracy. In smoke-only or low-light conditions, AHE-YOLO demonstrates higher sensitivity to reduced false activation in nonfire regions.

As shown in Figure 13, the heatmap comparisons visually confirm that the integration of the ADown, HS-FPN, and EMBC modules significantly enhances the attention of the model to critical fire-related regions, thereby improving both target perception accuracy and robustness across complex natural environments.

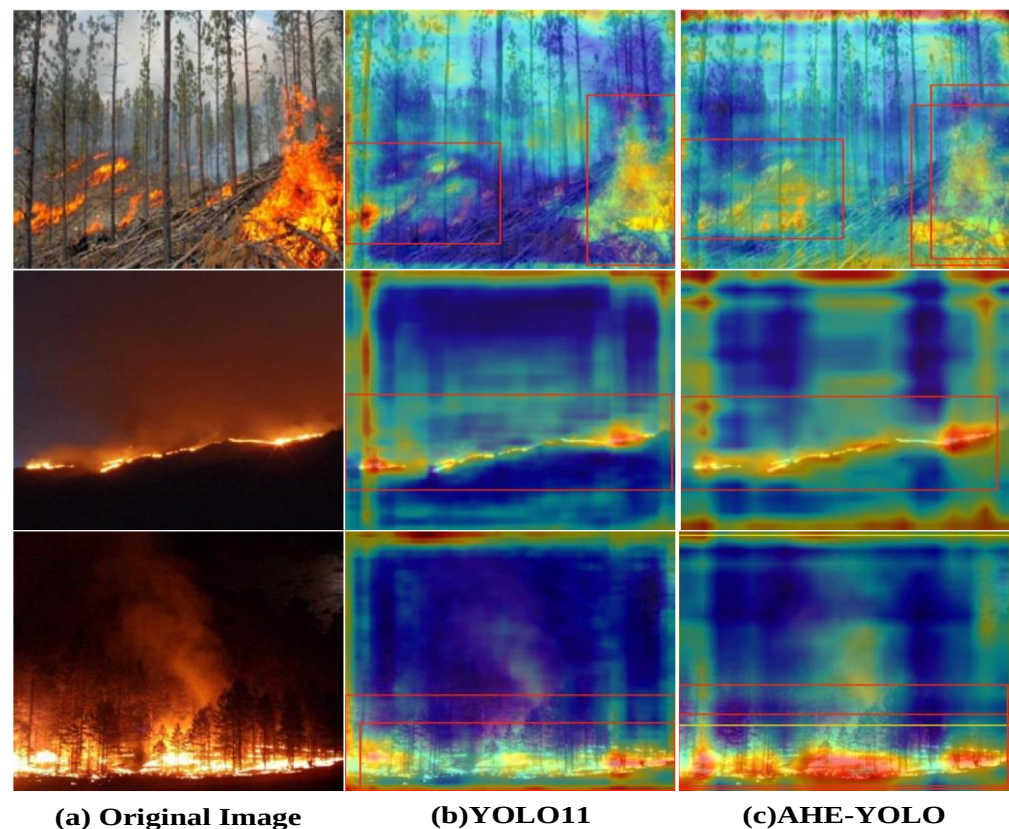


Figure 13. The heat map visualization comparison. Note: The color map represents the attention intensity, and the red boxes/lines highlight the regions of interest.

4.2. Limitations and Failure Analysis

Although the proposed AHE-YOLO achieves strong performance on the constructed dataset, certain limitations remain in challenging real-world scenarios. In particular, false positives may still occur under complex environmental conditions such as sunset illumination, haze, dense smoke dispersion, and atmospheric scattering. These factors can produce visual patterns that partially resemble fire or smoke, thereby increasing the risk of misclassification. This observation is generally consistent with previous studies on wildfire detection under complex atmospheric conditions [36].

In addition, extremely small or low-contrast early-stage smoke targets remain difficult to distinguish from background clutter at long distances. Although the proposed modules improve small-target sensitivity, the performance may still degrade in cases with severe occlusion or very weak smoke signatures.

Future work will focus on incorporating more hard negative samples, improving domain diversity, and exploring multi-modal information (e.g., thermal cues) to further enhance robustness and reduce false alarms in challenging environments.

5. Conclusions

This paper presents an improved lightweight forest fire detection algorithm, AHE-YOLO, based on YOLO11, which achieves a balanced trade-off between detection accuracy and computational efficiency for intelligent systems. The model comprises three core modules: the ADown downsampling optimization module, the HS-FPN multi-scale feature fusion module, and the EMBC effective feature enhancement module. These modules significantly reduce the complexity of the model and the requirements of computational resources while preserving high detection accuracy.

Experimental results on a self-built forest fire image dataset show that YOLO-AHE achieves a mAP of 94.8% and a recall of 85.1%, representing a 2.7% improvement in detection accuracy and a 16.5% increase in inference speed compared with YOLO11. At the same time, the number of model parameters is reduced by 39.7%, FLOPs are decreased by 27%, and model size is reduced by 36.4%. These results demonstrate that the proposed modules effectively enhance both model performance and resource efficiency.

This study demonstrates that incorporating efficient attention and feature fusion mechanisms significantly enhances UAV-based fire detection, which can be extended to other real-time remote sensing tasks, such as disaster monitoring and environmental surveillance.

The proposed lightweight architecture reduces computational burden and energy demand, which is particularly important for long-duration UAV missions and supports sustainable monitoring systems operating in remote forest regions.

In future work, we will focus on further improving the inference speed of the model and optimizing its deployment on resource-constrained platforms, such as UAVs and remote sensing systems. The stability, safety, and cross-regional generalization capabilities of the model under complex scenarios will be rigorously assessed, providing a foundation for the development of a robust and intelligent forest fire early warning system.

Author Contributions: Conceptualization, S.M. and Y.H.; methodology, Y.H.; validation, S.M., Y.H. and Y.W.; formal analysis, Y.H.; investigation, Y.H.; resources, S.M.; data curation, S.M.; writing—original draft preparation, Y.H.; writing—review and editing, S.M. and Y.H.; supervision, Y.Z. and Y.W.; project administration, Y.Z.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, Project No. 62103309.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is provided within the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

UAV	Unmanned Aerial Vehicle
SVM	Support Vector Machine
YOLO	You Only Look Once
mAP	mean Average Precision
Adown	Adaptive Downsampling
HS-FPN	High-level Screening-feature Fusion Pyramid Network
EMBC	Efficient Mobile Inverted Bottleneck Convolution
MBConv	Mobile Inverted Bottleneck Convolution
CA	Channel Attention
SSF	Selective Feature Fusion
DM	Dimension Matching
DSC	Depthwise Separable Convolution
MB	Model Size
SE	Squeeze-and-Excitation
SGD	Stochastic Gradient Descent
PR	Precision-Recall
AP	Average Precision
FLOPs	Floating Point Operations

FPS Frames Per Second

References

- Chen, Y.J.; Lai, Y.S.; Lin, Y.H. BIM-based augmented reality inspection and maintenance of fire safety equipment. *Autom. Constr.* **2020**, *110*, 103041. [\[CrossRef\]](#)
- Chuvieco, E.; Aguado, I.; Salas, J.; García, M.; Yebra, M.; Oliva, P. Satellite remote sensing contributions to wildland fire science and management. *Curr. For. Rep.* **2020**, *6*, 81–96. [\[CrossRef\]](#)
- Qiao, Y.; Jiang, W.; Wang, F.; Su, G.; Li, X.; Jiang, J. FireFormer: An efficient transformer to identify forest fire from surveillance cameras. *Int. J. Wildland Fire* **2023**, *32*, 1364–1380. [\[CrossRef\]](#)
- Kim, S.Y.; Muminov, A. Forest fire smoke detection based on deep learning approaches and unmanned aerial vehicle images. *Sensors* **2023**, *23*, 5702. [\[CrossRef\]](#)
- Yang, H.; Wang, J.; Wang, J. Efficient detection of forest fire smoke in UAV aerial imagery based on an improved YOLOv5 model and transfer learning. *Remote Sens.* **2023**, *15*, 5527. [\[CrossRef\]](#)
- Sudhakar, S.; Vijayakumar, V.; Kumar, C.S.; Priya, V.; Ravi, L.; Subramaniaswamy, V. Unmanned aerial vehicle (UAV)-based forest fire detection and monitoring for reducing false alarms. *Comput. Commun.* **2020**, *149*, 1–16. [\[CrossRef\]](#)
- Vasconcelos, R.N.; Rocha, W.J.S.F.; Costa, D.P.; Duverger, S.G.; de Santana, M.M.M.; Cambui, E.C.B.; Ferreira-Ferreira, J.; Oliveira, M.; da Silva Barbosa, L.; Cordeiro, C.L. Fire detection with deep learning: A comprehensive review. *Land* **2024**, *13*, 1696. [\[CrossRef\]](#)
- Harkat, H.; Ahmed, H.F.T.; Nascimento, J.M.P.; Bernardino, A. Early fire detection using wavelet-based features. *Measurement* **2025**, *242*, 115881. [\[CrossRef\]](#)
- Strässle, R.M.; Faldella, F.; Doll, U. Deep learning-based image segmentation for instantaneous flame front extraction. *Exp. Fluids* **2024**, *65*, 94. [\[CrossRef\]](#)
- Seydi, S.T.; Saeidi, V.; Kalantar, B.; Ueda, N.; Halin, A.A. Fire-Net: A deep learning framework for active forest fire detection. *J. Sens.* **2022**, *2022*, 8044390. [\[CrossRef\]](#)
- Terven, J.; Córdova-Esparza, D.M.; Romero-González, J.A. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [\[CrossRef\]](#)
- Zeng, G.; Zhang, X.; Wang, Z.; Xu, H.; Chen, Z.; Li, B.; Tu, Z. YOLO-Count: Differentiable object counting for text-to-image generation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Honolulu, HI, USA, 19–23 October 2025; pp. 16765–16775.
- Shen, D.; Chen, X.; Nguyen, M.; Yan, W.Q. Flame detection using deep learning. In Proceedings of the 4th International Conference on Control, Automation and Robotics (ICCAR), Auckland, New Zealand, 20–23 April 2018; pp. 416–420.
- Talaat, F.M.; ZainEldin, H. An improved fire detection approach based on YOLOv8 for smart cities. *Neural Comput. Appl.* **2023**, *35*, 20939–20954. [\[CrossRef\]](#)
- Li, J.; Tang, H.; Li, X.; Dou, H.; Li, R. LEF-YOLO: A lightweight method for intelligent detection of four extreme wildfires based on the YOLO framework. *Int. J. Wildland Fire* **2023**, *33*, 1. [\[CrossRef\]](#)
- Chen, G.; Zhang, Y.; Liu, X.; Jiao, W.; Bai, D.; Lin, H. LMDFS: A lightweight model for detecting forest fire smoke in UAV images based on YOLOv7. *Remote Sens.* **2023**, *15*, 3790. [\[CrossRef\]](#)
- Nie, Z.; Xu, Y.; Zhao, J.; Yuan, M. Fire classification and detection in imbalanced remote sensing images using a three-sphere model combined with YOLOv5. *Appl. Soft Comput.* **2025**, *113*, 192. [\[CrossRef\]](#)
- Gonnelli, A.; Baronti, S.; Carlà, R.; Raimondi, V. The impact of spatial resolution on active fire monitoring using multispectral satellite imagery. *Eng. Proc.* **2023**, *51*, 30.
- Niu, K.; Wang, C.; Xu, J.; Liang, J.; Zhou, X.; Wen, K.; Lu, M.; Yang, C. Early forest fire detection with UAV image fusion using visible and infrared sensors. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, early access. [\[CrossRef\]](#)
- Jiang, C.; Ren, H.; Ye, X.; Zhu, J.; Zeng, H.; Nan, Y.; Sun, M.; Ren, X.; Huo, H. Object detection from UAV thermal infrared images and videos using YOLO models. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102912. [\[CrossRef\]](#)
- Oliver, J.A.; Pivot, F.C.; Tan, Q.; Cantin, A.S.; Wooster, M.J.; Johnston, J.M. A machine learning approach to waterbody segmentation in thermal infrared imagery for wildfire mapping. *Remote Sens.* **2022**, *14*, 2262. [\[CrossRef\]](#)
- Almeida, J.S.; Huang, C.; Nogueira, F.G.; Bhatia, S.; de Albuquerque, V.H.C. EdgeFireSmoke: A novel lightweight CNN model for real-time video fire–smoke detection. *IEEE Trans. Ind. Inform.* **2022**, *18*, 7889–7898. [\[CrossRef\]](#)
- Geng, X.; Su, Y.; Cao, X.; Li, H.; Liu, L. YOLOFM: An improved fire and smoke object detection algorithm based on YOLOv5n. *Sci. Rep.* **2024**, *14*, 4543. [\[CrossRef\]](#) [\[PubMed\]](#)
- Casbeer, D.; Li, S.-M.; Beard, R.; Mehra, R.; McLain, T. Forest fire monitoring with multiple small UAVs. In Proceedings of the American Control Conference (ACC), Portland, OR, USA, 8–10 June 2005; pp. 3530–3535.
- Aral, R.A.; Zalluhoglu, C.; Akcapinar Sezer, E. Lightweight and attention-based CNN architecture for wildfire detection using UAV vision data. *Int. J. Remote Sens.* **2023**, *44*, 5768–5787. [\[CrossRef\]](#)

26. Wang, J.; Wang, Y.; Liu, L.; Yin, H.; Ye, N.; Xu, C. Weakly supervised forest fire segmentation in UAV imagery based on foreground-aware pooling and context-aware loss. *Remote Sens.* **2023**, *15*, 3606. [[CrossRef](#)]
27. Zhang, J.; Zhang, G.; Kang, Y.; Dong, Y.; Liu, Y.; Xie, S.; Xu, H. Determination of forest fire intensity level using multi-temporal satellite remote sensing imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, early access. [[CrossRef](#)]
28. Chen, C.; Zhou, Y.; Xie, W.; Meng, T.; Zhao, X.; Pang, Z.; Chen, Q.; Liu, D.; Wang, R.; Yang, V.; et al. Lightweight, thermally insulating, fire-proof graphite–cellulose foam. *Adv. Funct. Mater.* **2023**, *33*, 2204219. [[CrossRef](#)]
29. Zhai, N.; Tian, Y.; Wang, L.; Yao, J. DBFFNet: A dual-branch feature fusion framework for high-precision flame and smoke detection. *IEEE Access* **2025**, *13*, 179966–179982. [[CrossRef](#)]
30. Liu, R.M.; Su, W.H. APHS-YOLO: A lightweight model for real-time detection and classification of *Stropharia rugoso-annulata*. *Foods* **2024**, *13*, 1710. [[CrossRef](#)]
31. Zhao, L.; Wang, L. A new lightweight network based on MobileNetV3. *KSII Trans. Internet Inf. Syst.* **2022**, *16*, 1–15. [[CrossRef](#)]
32. Liu, W.; Chen, Q. GSAC-YOLO: Global-selective attention and convolutional networks for transmission tower detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, early access. [[CrossRef](#)]
33. Liu, L.; Huang, F.; Gu, L.; Fang, Y. Improved YOLOv7 based on reparameterization and attention mechanism for slab detection. *IEEE Trans. Instrum. Meas.* **2025**, early access.
34. Chen, F.; Yang, M.; Wang, Y. Recognition of Forest Fire Smoke Based on Improved YOLOv8n Model. *Fire Technol.* **2025**, *61*, 3351–3374. [[CrossRef](#)]
35. Chen, X.; Zhu, E.; Zhu, Y.; Fu, Y.; Han, K.; Liu, B. WA-YOLO: A forest fire smoke detection method based on wavelet convolution and SimAM. *J. Supercomput.* **2025**, *81*, 1517. [[CrossRef](#)]
36. Xu, R.; Lin, H.; Lu, K.; Cao, L.; Liu, Y. A forest fire detection system based on ensemble learning. *Forests* **2021**, *12*, 217. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.