

Article

TriRHC-YOLO: A Method for Early Forest Fire Detection in Complex Environments Based on UAV Images

Bo Song^{1,2}, Bo Li^{1,2}, Zhiyong Zhang^{1,2}, Yun Chen^{1,2,*}, Qingyang Wang¹, Xing Zhang¹, Zhen Cao¹, Tao Yue^{1,2} and Jianwu Jiang¹

¹ College of Geomatics and Geoinformation, Guilin University of Technology, Guilin 541004, China; bosong@glut.edu.cn (B.S.); 2120242157@glut.edu.cn (B.L.); 2120242221@glut.edu.cn (Z.Z.); qingyang@glut.edu.cn (Q.W.); zhangxing@glut.edu.cn (X.Z.); caozhen@glut.edu.cn (Z.C.); yuetao@glut.edu.cn (T.Y.); fengbuxi@glut.edu.cn (J.J.)

² Guangxi Key Laboratory of Ecological Spatiotemporal Big Data Perception Service, Guilin 541004, China

* Correspondence: chenyun@glut.edu.cn

Abstract

To address the problems of small fire-spot scale, blurred boundaries, complex backgrounds, and insufficient feature representation of weak targets in Unmanned Aerial Vehicle (UAV)-based early forest fire detection, a YOLOv8n-based forest fire detection model, termed TriRHC-YOLO, is proposed. The model first introduces Reparameterized VGG (RepVGG)Block into the backbone network to enhance the extraction capability of shallow local features. Subsequently, a Hierarchical Feature Attention (HFA) module is designed to collaboratively model fire-spot features from three levels, namely directional structures, local textures, and global semantics, thereby enhancing the network's capability to discriminate fire targets and suppressing interference from complex forest backgrounds. Finally, a Cross Stage Partial Feature Fusion with Cascade Star Block (C2f-CStar) module is designed to improve the representation capability of the model for local structural information and weak salient fire-spot features under complex backgrounds through cascaded spatial feature reconstruction and a star-shaped multiplicative gating mechanism. In addition, a UAV-specific early forest fire detection dataset is constructed based on the FLAME and FLAME_VISION datasets, and experimental validation is conducted on this dataset. The experimental results show that the proposed TriRHC-YOLO outperforms several classical YOLO algorithms, including YOLO11n, YOLO12, and YOLO26, as well as six advanced YOLO-based improved models. The Recall, mean Average Precision (mAP)@0.5, and mAP@0.5:0.95 reach 0.769, 0.848, and 0.608, respectively. The results of the ablation experiments further verify the effectiveness of the three designed modules. Moreover, the proposed model contains only 3.181 M parameters and achieves 168.251 Frames Per Second (FPS), demonstrating favorable real-time detection capability. Overall, the proposed method can effectively improve the detection accuracy of early weak fire targets and the background suppression capability under complex forest backgrounds, making it suitable for real-time UAV-based forest fire inspection tasks.



Academic Editor: Kevin Tansey

Received: 2 June 2026

Revised: 29 July 2026

Accepted: 2 August 2026

Published: 6 August 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.

Keywords: forest fire detection; UAV imagery; YOLOv8n; small object detection; attention mechanism

1. Introduction

Forest fires are major disasters that threaten the stability of forest ecosystems and regional public safety, and are characterized by concealment, sudden onset, and rapid

spread [1,2]. Under meteorological conditions such as high temperature, drought, and strong winds in particular, small-scale incipient ignition points may rapidly develop into large fire fronts, imposing substantial pressure on fire assessment, personnel deployment, and on-site suppression operations [3–5]. In their early stages, forest fires typically manifest as small-scale flames, localized smoke, weak thermal anomalies, and low-salience visual characteristics. Accurate detection of early-stage forest fires can substantially reduce the probability of fire spread, alleviate suppression difficulty and resource consumption, and maximize the protection of forest ecosystems and the safety of surrounding residents [6–8]. Therefore, improving the detection capability for weak early-stage forest-fire targets in complex forest backgrounds is of great significance for constructing intelligent forest fire prevention systems, enhancing the efficiency of unmanned patrol inspection, and strengthening the timeliness of emergency responses.

In recent years, UAVs have increasingly become important platforms for forest-fire patrol and inspection, owing to their high maneuverability, controllable viewing perspectives, relatively low deployment costs, and strong adaptability to complex terrain [9–11]. Early studies on fire-spot recognition in UAV imagery primarily relied on conventional image-processing techniques and hand-crafted features [12–15]. For example, color-based methods typically perform threshold segmentation by exploiting the high brightness and high saturation of flames in Red, Green, Blue (RGB) or Hue, Saturation, Value (HSV) color spaces [16], whereas texture- and motion-based methods identify targets by analyzing the dynamic variations in flame regions or local texture differences [17]. In addition, several studies have combined edge detection, region growing, and related techniques to extract fire-spot regions [18–20]. Although these methods can be effective in relatively simple scenarios, their performance is highly dependent on empirical thresholds and manually designed features. Consequently, they exhibit limited adaptability to illumination variations, smoke interference, and complex backgrounds, making it difficult to achieve stable detection performance in real-world forest environments.

Concurrently, advances in deep-learning-based object detection have provided a new technical pathway for the automatic recognition of forest fires in UAV imagery, with convolutional neural networks (CNNs) increasingly being introduced into fire detection tasks [21–25]. Among existing deep learning approaches, object detection models such as Faster R-CNN, DETR, and the YOLO series have been widely applied to the recognition of flames, smoke, and forest-fire targets. In particular, YOLO-family models offer end-to-end detection, fast inference, and convenient deployment, making them well suited to time-critical applications such as UAV-based patrol inspection.

Frizzi et al. [26] were among the first to apply CNNs to fire detection. By leveraging the capability of CNNs to automatically learn discriminative features, their method effectively overcame several limitations of conventional approaches and achieved promising performance. From the perspective of feature fusion, Li et al. [27] improved the neck architecture of the detection network and proposed the YOLOGX algorithm, which enhanced the detection performance for multiscale fire targets in complex terrain; however, false detections may still occur under interference from complex backgrounds. Mingxing Tan et al. [28] first proposed the weighted bidirectional feature pyramid network (BiFPN), a core component of the EfficientDet detector, to strengthen multiscale feature fusion by introducing learnable weights that evaluate the relative importance of different input features. Meanwhile, to reduce false alarms under illumination variations and in environments containing visually similar distractors, attention mechanisms have been widely incorporated into fire detection models. For example, Yaxuan Pang et al. [29] introduced Efficient Channel Attention (ECA) and SimAM, a parameter-free spatial attention mechanism, to enhance fire-spot localization cues along both channel and spatial dimensions. By adaptively fusing

multiple attention-enhanced features, their method improved the perception of small targets. Hikmat Yar et al. [30] designed a dual fire-attention mechanism combining channel attention (CA) and spatial attention (SA), and proposed a hybrid CNN–InceptionV3 architecture capable of more comprehensively capturing the visual characteristics of fire, thereby improving fire detection performance. By strengthening key features and suppressing redundant background responses, these methods optimized fire recognition in complex scenes; nevertheless, their performance remains susceptible to domain shifts between application scenarios and training samples. At the model-architecture level, research has primarily focused on enhancing network sensitivity to tiny and early-stage targets. For instance, Wanhong Yang et al. [31] proposed SIMCB-YOLO, which expands the receptive field of the model, strengthens smoke-feature extraction, and improves small-object detection accuracy under smoke interference. To meet the real-time requirements of airborne detection, Dilshad et al. [32] embedded a lightweight joint attention module to enhance flame and smoke feature representations, maintaining an extremely compact model size with only a slight increase in computational overhead, thereby enabling real-time fire detection deployment on edge devices. Gao et al. [33] introduced the RepVGG module into the YOLO framework. With multi-branch structure for training and single-branch structure for inference, this method improves model feature representation and reduces model complexity.

YOLOv8n provides a strong foundation in terms of detection accuracy, architectural design, and engineering deployment. However, when directly applied to early-stage forest fire detection from UAV imagery, it still faces several limitations [34]. First, in high-altitude or oblique UAV views, incipient fire spots are typically small in scale and occupy only a limited proportion of image pixels; consequently, shallow local features are easily weakened during repeated downsampling. Second, natural backgrounds such as forests, grasslands, bare soil, and croplands exhibit highly complex textures, while flame colors may resemble bright background regions, dry branches and fallen leaves, or shadow boundaries, thereby increasing the risk of false detections. Third, early-stage flames often have blurred boundaries and unstable smoke-transition regions, whereas the original network is insufficient in collaboratively modeling directional contours, fine-grained textures, and high-level semantic information. Fourth, the discriminative features of weak targets in complex scenes remain inadequate, making it difficult for the model to simultaneously achieve high recall, accurate localization, and real-time detection.

Based on the aforementioned issues, this paper proposes the TriRHC-YOLO model for UAV-based forest fire detection. The model is built upon YOLOv8n as the baseline and specifically addresses the problems of insufficient structured representation of early weak fire-spot targets, strong background interference, and missing discriminative features under complex forest backgrounds through targeted improvements:

(1) RepVGGBlock is introduced into the backbone network to enhance the stable extraction capability of shallow local features, such as flame edges and smoke transition regions.

(2) The HFA mechanism is designed to collaboratively enhance key discriminative information from three levels—direction, texture, and semantics—while suppressing interference from complex forest backgrounds.

(3) The C2f-CStar module is constructed to improve the representation capability of local structural information and weak fire-spot features under complex backgrounds through cascaded structural reconstruction and a star-shaped multiplicative gating mechanism.

These three types of modules jointly form a progressive enhancement path with stable shallow local features, mid-level structural reconstruction, and high-level attribute collaboration, enabling the model to better meet the scene requirements of UAV-based early forest fire detection where small targets, weak textures, blurred boundaries, and complex backgrounds coexist.

2. Materials and Methods

To address the challenges of insufficient structured representation, susceptibility to background interference, and lack of distinguishing features in the detection of weak forest fire targets by UAVs in complex forest environments, this study builds upon YOLOv8n and performs hierarchical correction and progressive enhancement targeting the representation defects of flame targets at different levels, proposing an improved network for UAV-based weak-fire-spot detection: TriRHC-YOLO. The model architecture is shown in Figure 1.

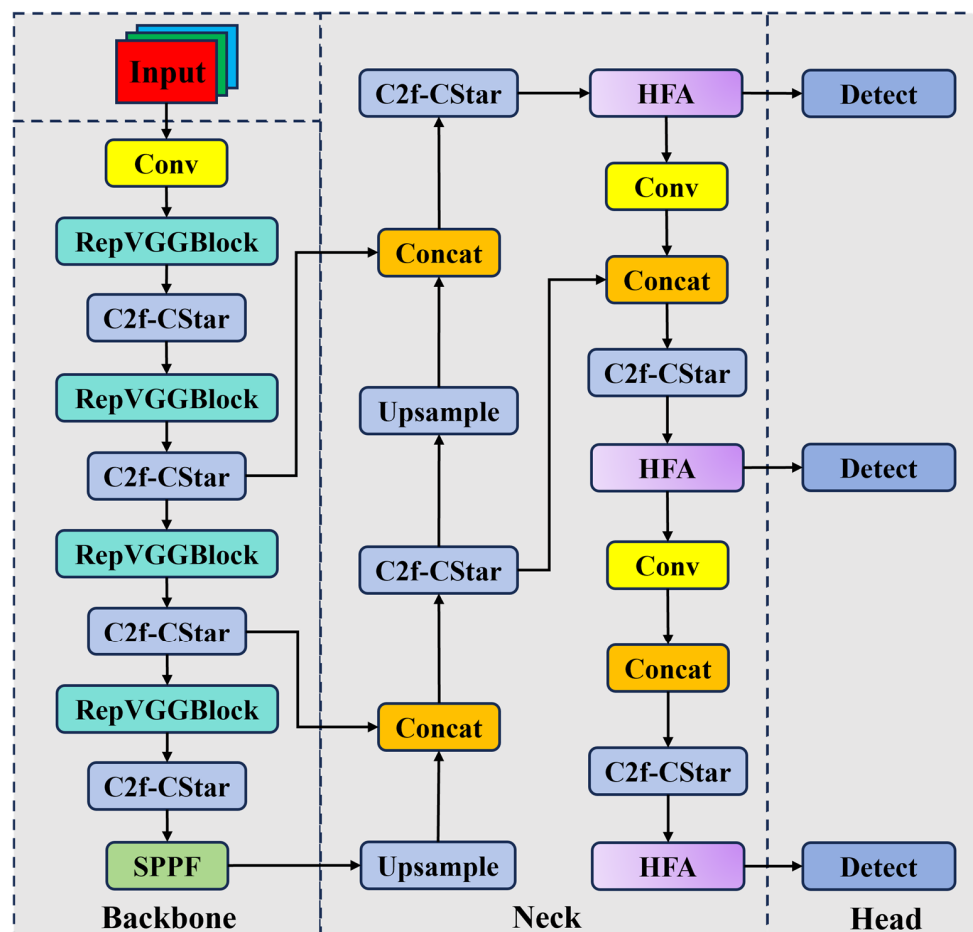


Figure 1. Structure Diagram of the TriRHC-YOLO Model.

The overall design performs hierarchical correction targeting the representation defects of flame targets at different levels: first, RepVGGBlock is introduced into the backbone, leveraging multi-branch convolutions during training to enhance local feature extraction capability, and folded into a single-branch convolution via structural reparameterization during inference, thereby enhancing the stable extraction of shallow features such as flame edges and smoke transition regions without significantly increasing deployment burden; second, at the high-level feature integration stage, the HFA is designed to decouple and model three types of complementary information—directional contours, local textures, and global semantics—and achieve adaptive fusion through learnable weights; finally, to address the deficiencies of the original C2f in modeling spatial features under complex scenarios, limited receptive field, and insufficient weakly salient channel collaboration, the C2f-CStar module is constructed. Through cascaded feature splitting, multiplicative channel gating, and spatial context closure, it progressively reconstructs internal target structural information, improving the model’s discriminative capability for early weak flame targets and background suppression ability.

These three modules together form a progressive enhancement path with stable shallow local features, mid-level structural reconstruction, and high-level attribute collaboration, enabling the network to maintain favorable real-time performance while better adapting to scenarios in UAV-based forest fire detection where weak targets, blurred boundaries, and complex backgrounds coexist.

2.1. RepVGG Module

To address the issues of insufficient shallow structural representation caused by blurred boundaries, unstable local features, and strong background interference of weak early-stage forest fire targets in complex natural environments, inspired by the work of Gao et al. [33], this study introduces the Reparameterized VGG (RepVGG) module into the backbone network of YOLOv8n to replace several original basic feature extraction units. This modification enhances the model's capability to extract discriminative features from flame and smoke targets while reducing the number of model parameters. The core idea of RepVGG is to employ a multi-branch convolutional architecture during the training stage to improve representation capability while converting the multi-branch structure into an equivalent single 3×3 convolutional structure through structural reparameterization during the inference stage. This design reduces the complexity of the inference graph and improves deployment efficiency while preserving feature learning capability [35]. The module architecture during the training stage is shown in Figure 2, where Figure 2a illustrates the conventional feature extraction module and Figure 2b presents the module structure adopted in the downsampling stage.

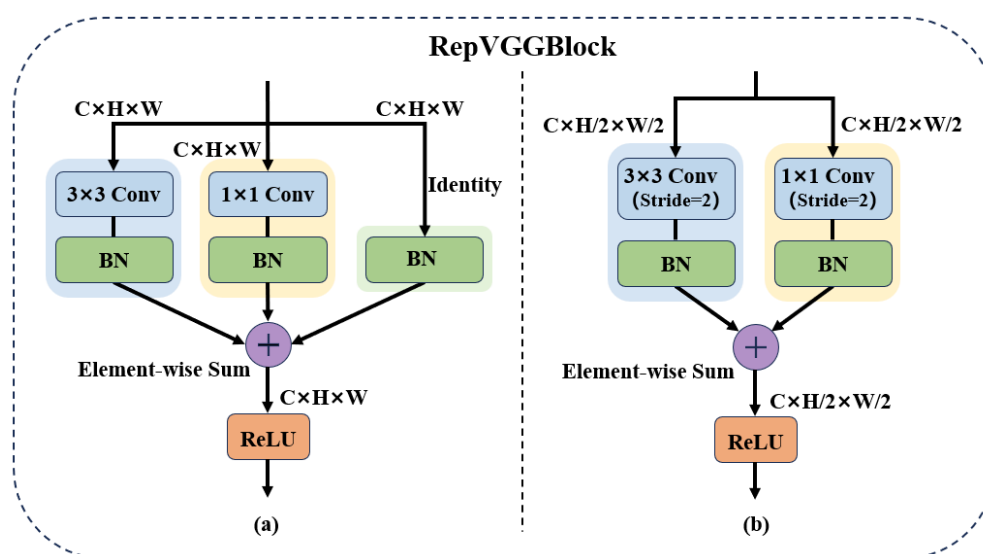


Figure 2. Structure Diagram of the RepVGG Block. (a) Conventional feature extraction module; (b) Module structure used in the downsampling stage.

In terms of its implementation mechanism, the key to RepVGG lies in the seamless integration of convolutional layers with Batch Normalization (BN) layers. For the i -th channel, the BN transformation can be expressed as:

$$y_i = \frac{x_i - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}} \gamma_i + \beta_i \quad (1)$$

where μ_i and σ_i denote BN statistics, γ_i and β_i are learnable parameters, and ϵ is a numerical stabilizer. Furthermore, the convolutional kernel parameters can be rewritten as:

$$W'_i = \frac{\gamma_i}{\sigma_i} W_i, b'_i = \beta_i - \frac{\mu_i \gamma_i}{\sigma_i} \quad (2)$$

where W'_i and b'_i denote the equivalent convolution kernel and bias after BN fusion, respectively. Based on the above transformation, the multi-branch architecture employed during training can be re-parameterized into a single convolutional layer during inference, thereby reducing the parameter access overhead and computational cost introduced by additional branches.

2.2. HFA Module

To address the problems of target directional features being easily lost, insufficient representation of fine-grained texture features, and uneven channel semantic responses in forest fire-spot recognition, this paper proposes a Hierarchical Feature Attention (HFA) module. The architecture of the module is illustrated in Figure 3. The module collaboratively models fire-spot features from three levels, namely directional structures, local textures, and global semantics. Residual connections are employed to enhance feature representations, thereby improving the network's feature representation and discriminative capability for key targets while preserving the original discriminative information.

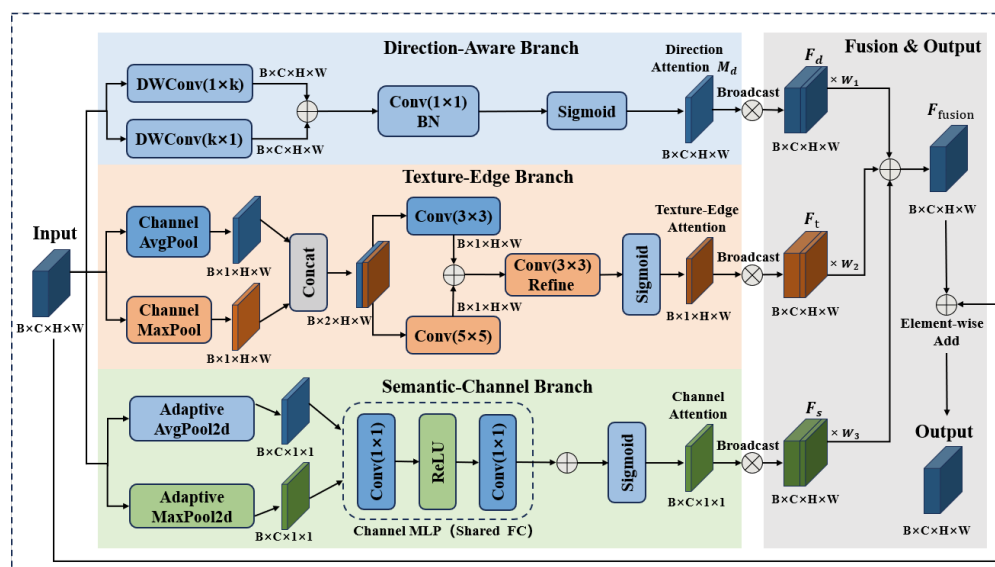


Figure 3. Structure Diagram of the HFA Block. (The “+” symbol denotes feature fusion/addition at different stages, and the “×” symbol represents element-wise multiplication).

Let the input feature map be denoted as X , where B, C, H and W represent the batch size, channel number, feature height, and feature width, respectively. The three branches independently model the input feature X to obtain the direction-enhanced feature F_d , texture-enhanced feature F_t , and semantic-enhanced feature F_s . Subsequently, F_d , F_t and F_s are adaptively fused through learnable scalar weights, followed by feature aggregation using a 1×1 convolution and BN. Finally, a residual connection is employed to add the fused representation to the input feature, yielding the enhanced output feature Y . The overall process can be formulated as follows:

$$\begin{aligned} F_{fusion} &= w_1 F_d + w_2 F_t + w_3 F_s \\ F_{out} &= BN\left(Conv_{1 \times 1}(F_{fusion})\right) \\ Y &= X + F_{out} \end{aligned} \quad (3)$$

where w_1, w_2 and w_3 are learnable fusion parameters.

The first branch is primarily designed to model the structural continuity and spatial distribution characteristics of fire-spot targets. By enhancing horizontal and vertical structural responses, it strengthens the network's ability to focus on flame contours, edge-continuous regions, and elongated target regions, thereby improving the salient representation of fire-spot regions under complex backgrounds. Specifically, the input feature X is processed by a $1 \times k$ depthwise convolution and a $k \times 1$ depthwise convolution to extract horizontal and vertical directional features, respectively. The extracted features are then fused and further integrated through a 1×1 convolution followed by BN. Finally, a Sigmoid activation function is applied to generate the directional attention map. This process can be formulated as follows:

$$M_d = \sigma(\text{BN}(\text{Conv}_{1 \times 1}(\text{DWConv}_{1 \times k}(X) + \text{DWConv}_{k \times 1}(X)))) \quad (4)$$

where DWConv denotes the depthwise convolution operation and $\sigma(\cdot)$ represents the Sigmoid activation function. The generated directional attention map is then multiplied elementwise with the input feature to obtain the direction-enhanced feature:

$$F_d = X \odot M_d \quad (5)$$

where $M_d \in \mathbb{R}^{B \times C \times H \times W}$. Therefore, the output feature F_d preserves the same spatial dimensions as the input feature.

The second branch is primarily used to supplement local textures, edge contours, and fine-grained spatial details of fire-spot targets, aiming to improve the model's ability to recognize key local regions under complex backgrounds. Specifically, average pooling and max pooling are first applied to the input feature X along the channel dimension to generate two single-channel spatial descriptor maps. These two descriptors are subsequently concatenated along the channel dimension to form a two-dimensional statistical representation. Next, parallel 3×3 and 5×5 convolutions are employed to perform spatial modeling on S . The extracted features are then refined through a 3×3 convolution to generate the texture–edge attention map. This process can be formulated as follows:

$$\begin{aligned} A &= \text{AvgPool}_c(X) \\ M &= \text{MaxPool}_c(X) \\ M_t &= \sigma(\text{Conv}_{3 \times 3}(\text{Conv}_{3 \times 3}([A, M]) + \text{Conv}_{5 \times 5}([A, M]))) \end{aligned} \quad (6)$$

where $A, M, M_t \in \mathbb{R}^{B \times 1 \times H \times W}$. The attention map is subsequently expanded to all channels through a broadcasting operation and multiplied elementwise with the input feature to obtain the texture-enhanced feature:

$$F_t = X \odot M_t \quad (7)$$

The third branch is used to extract semantic information from the global context and adaptively adjust the response strength of different channels, thereby highlighting key semantic channels and suppressing background interference. Specifically, global average pooling and global max pooling are first applied to the input feature X to obtain two global semantic descriptor vectors: $z_{avg}, z_{max} \in \mathbb{R}^{B \times C \times 1 \times 1}$. The two descriptors are subsequently fed into a shared channel mapping network composed of two 1×1 convolution layers with a ReLU activation function inserted between them to achieve channel-wise dimensionality reduction and restoration. The process can be expressed as follows:

$$\phi(z) = W_2 \delta(W_1 z) \quad (8)$$

where W_1 and W_2 denote the weights of the two mapping layers, respectively, and $\delta(\cdot)$ represents the ReLU activation function. Finally, the obtained channel attention weights are broadcast along the spatial dimensions and multiplied elementwise with the input feature to generate the semantic-enhanced feature:

$$F_s = X \odot \sigma(\phi(z_{avg}) + \phi(z_{max})) \quad (9)$$

The above three-layer progressive enhancement structure, combined with residual connections and input feature fusion, achieves collaborative enhancement of directional, texture, and semantic features, effectively compensating for the insufficient feature representation of weak forest fire targets and significantly improving the network's feature representation and discriminative capability for fire-spot targets.

2.3. C2f-CStar Module

To address the problems of insufficient local spatial structure modeling capability, limited receptive field, and restricted intermediate-level feature reorganization capability of YOLOv8n in complex forest scenes, this paper proposes a Cross Stage Partial Feature Fusion with Cascade Star Block (C2f-CStar) feature extraction module for the progressive modeling and effective reconstruction of internal target structural information. The module replaces the original Bottleneck module with the proposed CStarBlock (Cascade Star Block). Through cascaded DWConv and residual connections, it achieves progressive receptive field expansion and effective capture of multi-directional features. Meanwhile, a Star Multiplicative Gating mechanism is introduced to adaptively weight channel features, significantly enhancing the model's representation capability for small targets and complex backgrounds. The architecture of the module is shown in Figure 4.

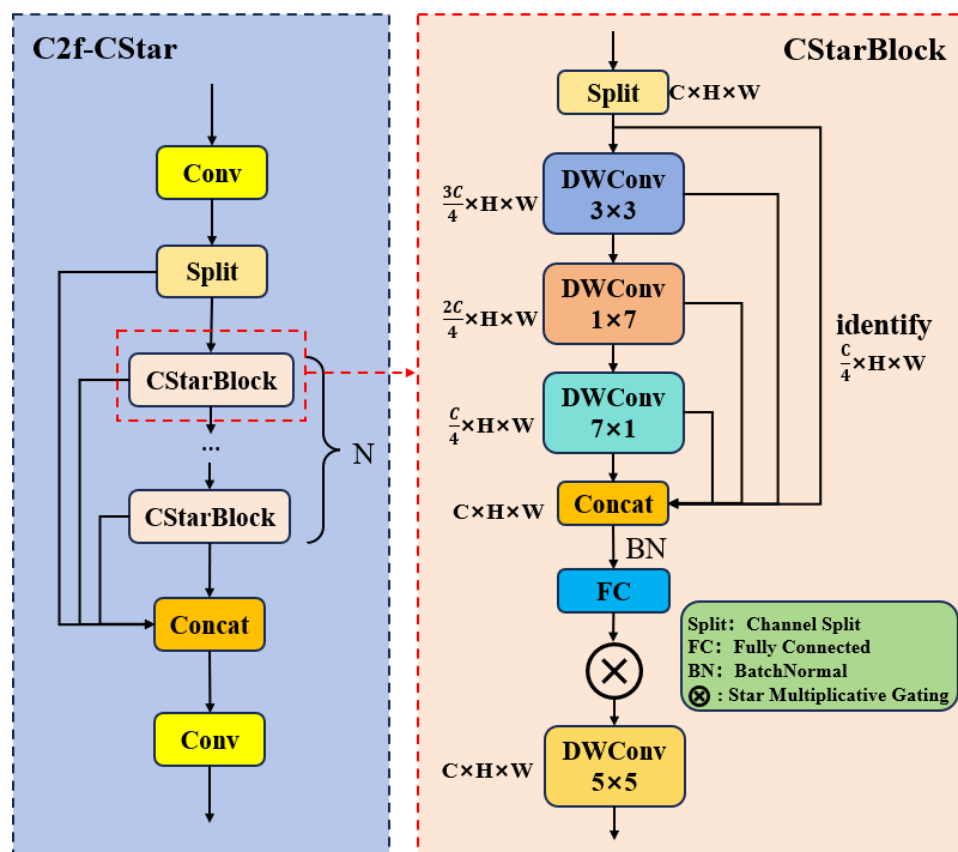


Figure 4. Structure Diagram of the C2f-CStar Block.

In the C2f-CStar module, let the input feature map be denoted as $X \in \mathbb{R}^{C \times H \times W}$, where C , H and W represent the channel number, feature height, and feature width, respectively. After passing through a 1×1 convolution, X is divided into two branches along the channel dimension through a Split operation. One branch is directly retained as the cross-stage residual pathway while the other branch sequentially passes through (N) CStarBlocks for hierarchical feature extraction. Finally, the outputs of the two branches are fused via a Concat operation, followed by a 1×1 convolution for channel compression before being forwarded to the next layer. This design not only preserves the gradient reuse advantage of the original C2f structure, but also introduces stronger feature representation capability through the proposed CStarBlock.

Within the CStarBlock, the input feature X is first split along the channel dimension into two sub-feature maps. Specifically, X_A serves as the identity branch to preserve the original feature information and participates in the subsequent Concat operation, whereas X_B is processed by a 3×3 DWConv to extract local structural details. Meanwhile, a residual connection is employed to maintain stable information flow:

$$X_{B1} = DWConv_{3 \times 3}(X_B) + X_B \in \mathbb{R}^{3C/4 \times H \times W} \quad (10)$$

Subsequently, X_{B1} is further split along the channel dimension. The first $C/4$ channels are projected through a 3×3 convolution and forwarded to the Concat pathway, while the remaining $2C/4$ channels are processed by a 1×7 DWConv for horizontal directional modeling. Similarly, the resulting feature X_{B2} is further divided, where the first $C/4$ channels are projected and concatenated, and the remaining $C/4$ channels are processed by a 7×1 DWConv for vertical directional modeling:

$$\begin{aligned} X_{B2} &= DWConv_{1 \times 7}(X_{B1,cont}) + B_{1,cont} \in \mathbb{R}^{2C/4 \times H \times W} \\ X_{B3} &= DWConv_{7 \times 1}(X_{B2,cont}) + X_{B2,cont} \in \mathbb{R}^{C/4 \times H \times W} \end{aligned} \quad (11)$$

Finally, the four branch features X_A , X_{B1} , X_{B2} and X_{B3} are concatenated along the channel dimension to restore the feature representation with channel dimension $C \times H \times W$, followed by BN for feature normalization:

$$F_{concat} = BN(Concat[X_A, X_{B1}, X_{B2}, X_{B3}]) \in \mathbb{R}^{C \times H \times W} \quad (12)$$

To further enhance inter-channel feature interaction, a Star Multiplicative Gating mechanism is introduced. Specifically, Global Average Pooling (GAP) is first applied to F_{concat} to extract channel-wise global statistical information. Subsequently, the pooled feature vector (z) is fed into two fully connected (FC) layers, denoted as FC1 and FC2, both of which expand the channel dimension to $(4C)$. A ReLU6 activation function is employed after FC1 to introduce nonlinear representation capability, whereas FC2 maintains a linear mapping:

$$z_1 = \text{ReLU6}(FC1(z)) \in \mathbb{R}^{4C}, z_2 = FC2(z) \in \mathbb{R}^{4C} \quad (13)$$

The outputs of the two branches are subsequently combined through elementwise multiplication to achieve implicit high-order feature interaction. The resulting feature is then reduced to the original channel dimension (C) through FC3 and expanded into a weight tensor of size $C \times 1 \times 1$. Finally, the generated weights are multiplied channelwise with F_{concat} to achieve adaptive feature recalibration:

$$w = FC3(z_1 \odot z_2) \in \mathbb{R}^C, F_{weighted} = F_{concat} \odot \text{Reshape}(w, [C, 1, 1]) \quad (14)$$

where $\odot$ denotes the Hadamard product (elementwise multiplication). The recalibrated feature map $F_{weighted}$ is subsequently processed by a 5×5 DWConv for spatial context fusion, thereby enlarging the receptive field and further capturing global contextual information. The final output feature map $Y \in \mathbb{R}^{C \times H \times W}$ can be expressed as:

$$Y = DWConv_{5 \times 5}(F_{weighted}) \quad (15)$$

In summary, C2f-CStar achieves multi-scale contextual modeling through a cascaded residual structure, enhancing the network's spatial feature extraction capability. At the same time, it leverages the Star Multiplicative Gating mechanism to strengthen the adaptive expression capability across channels, significantly improving the model's feature modeling ability for complex scenarios while maintaining a lightweight design.

3. Results

3.1. Dataset

To address the challenges of insufficient weak-target samples and limited scene coverage in early-stage forest fire detection under complex forest backgrounds, this study constructed a UAV-based early wildfire detection dataset by screening and integrating two publicly available datasets, FLAME [36] and FLAME_VISION [37]. The resulting dataset contains a total of 5000 UAV-acquired images of early-stage forest fires. Unlike conventional fire datasets, this dataset places greater emphasis on weak flame targets in natural environments, primarily covering incipient fire scenarios in forests, grasslands, fields, and other low-vegetation areas.

Specifically, during the selection process, the scale of fire targets was first categorized based on the annotated bounding box area A_b to characterize the distribution characteristics of fire targets at different scales in UAV images. Subsequently, samples were selected based on the fire target scale, with priority given to retaining small-scale early fire samples with a low proportion of flame regions in the image, blurred target boundaries, and weak visual saliency. Furthermore, considering occlusion and image quality factors, samples exhibiting smoke diffusion, canopy occlusion, and long-distance imaging blur were preferentially retained. Images containing large-scale combustion, non-UAV viewpoints, or overly simple backgrounds that did not conform to the characteristics of early forest fire detection were excluded, while images depicting complex low-vegetation environments such as forests, grasslands, and fields were preserved. Finally, all selected images were uniformly annotated in YOLO format, and the bounding box locations, category labels, and annotation completeness were carefully reviewed and corrected to ensure dataset quality.

The distribution of samples in the dataset is shown in Table 1. Overall, the overall distribution of various scenarios is relatively balanced and predominantly reflects the characteristics of early forest fires captured by UAVs: first, the shooting angles are mainly high-altitude top-down or oblique views, with target morphology affected by the viewing angle, as shown in Figure 5a; second, flame targets are small in scale, occupy a low proportion, and are sparsely embedded in complex backgrounds, as shown in Figure 5b; third, issues such as smoke diffusion, local shadows, and long-distance blur make structured representation of early flame targets difficult, as shown in Figure 5c; fourth, the background primarily consists of complex natural textures such as tree canopies, bare soil, grass, and farmland, which can easily be confused with flame color and edges, as shown in Figure 5d. Based on these characteristics, this dataset can effectively simulate the observational challenges of weak early forest fire targets under UAV inspection conditions, making it more suitable for studying the structured representation and detection methods of weak early forest fire targets.

Table 1. Statistics on the Distribution of Dataset Samples. (Since an image typically contains multiple features—such as complex backgrounds, smoke interference, and object occlusion—scene features are analyzed using multi-label statistics; therefore, the sum of the counts for each category exceeds the total number of samples in the dataset.)

Statistical Dimension	Subcategory	Number of Images (Pieces)	Percentage
Shooting Angles	Bird's-eye view	2993	59.86%
	Oblique view from high altitude	2007	40.41%
Fire Spot Scale	Small-scale fire points ($A_b \leq 32^2 \text{px}^2$)	3671	73.42%
	Medium-scale fire points ($32^2 < A_b \leq 96^2 \text{px}^2$)	1150	23.00%
	Large-scale fire points ($A_b > 96^2 \text{px}^2$)	179	3.58%
Scene Characteristics	Weak targets at high altitude and long range detected by UAVs	2766	55.32%
	Complex backgrounds such as tree canopies, bare ground, and farmland	4280	85.60%
	Scenes where flames transition into smoke	1824	36.48%
	Scenes obscured by tree canopies or smoke	3280	65.60%

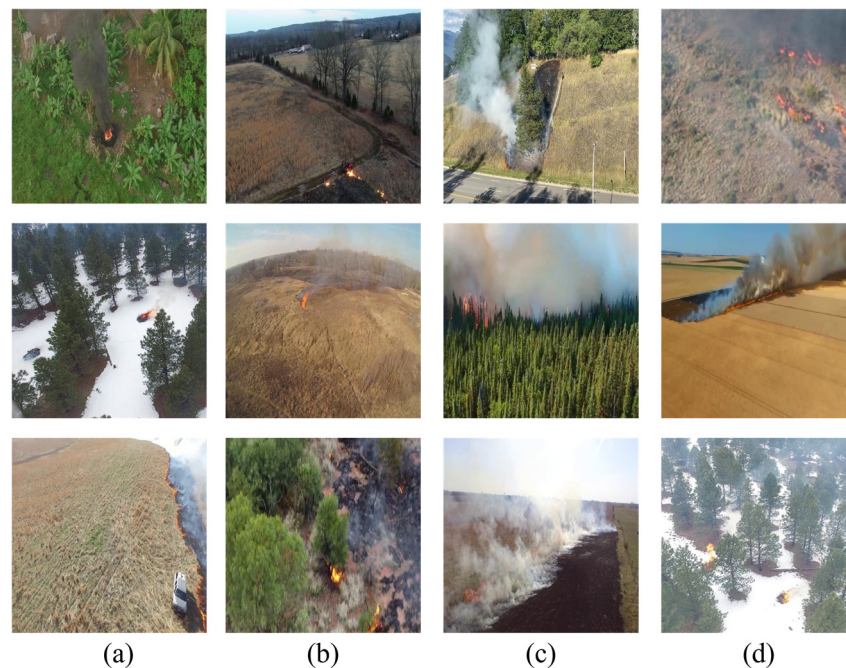


Figure 5. Early UAV-Based Wildfire Dataset. (a) Flame scenes captured from high-altitude top-down and oblique perspectives; (b) Scenes with small-scale sparse flame targets; (c) Scenes with smoke diffusion and shadows; (d) Scenes featuring occlusion and easily confused backgrounds.

3.2. Accuracy Evaluation Metrics

To comprehensively evaluate the performance of the proposed model in the UAV-based early forest fire weak-target detection task, Recall, mAP@0.5, and mAP@0.5:0.95 are adopted as the primary evaluation metrics. The model is comprehensively assessed from two aspects, namely target detection capability and detection accuracy. Among them, Recall is used to measure the capability of the model to correctly detect target objects, and is calculated as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (16)$$

where TP denotes the number of correctly detected fire spots, and FN denotes the number of missed fire spots. A higher Recall indicates fewer missed detections and a stronger capability of the model to identify weak early forest fire targets.

mAP@0.5 represents the mean Average Precision calculated under an Intersection over Union (IoU) threshold of 0.5. Specifically, a detection result is regarded as a successful match when the IoU between the predicted bounding box and the ground-truth bounding box exceeds 0.5. This metric reflects the detection capability of the model under relatively relaxed localization conditions, and is calculated as follows:

$$mAP50 = \frac{1}{N} \sum_{i=1}^N AP_i^{IoU=0.5} \quad (17)$$

where N denotes the number of categories, and $AP_i^{IoU=0.5}$ represents the Average Precision (AP) of the (i)-th category under the IoU threshold of 0.5.

mAP@0.5:0.95 represents the average of AP values calculated under multiple IoU thresholds ranging from 0.5 to 0.95 with a step size of 0.05. It can objectively reflect the overall localization performance of the model. The calculation formula is given as follows:

$$mAP50-95 = \frac{1}{10} \sum_{k=1}^{10} mAP@IoU_k, \quad (18)$$

$$IoU_k \in \{0.5, 0.55, \dots, 0.95\}$$

3.3. Test Environment and Parameter Settings

The experiments were conducted on the Ubuntu 20.04.5 LTS operating system. All model training and inference were implemented using the PyTorch 2.0.0 framework and executed on a single NVIDIA vGPU with 32 GB of GPU memory, ensuring a consistent experimental environment for all models. During the experiments, the dataset was randomly divided into a training set (3500 images), a validation set (500 images), and a test set (1000 images) according to a ratio of 7:1:2. Furthermore, to evaluate the real-time deployment capability of the proposed model on UAV edge devices, the trained models were deployed on an NVIDIA Jetson Orin NX embedded platform for inference speed evaluation. The frame rate was measured to assess the real-time detection performance of the model in practical UAV applications. No pretrained weights were used in any of the experiments. The detailed experimental settings are summarized in Table 2.

Table 2. Experimental Parameters.

Parameter Name	Value
Number of training epochs	300
Batch size	32
Input image dimensions	640 × 640
Optimizer	SGD
Learning rate	0.01
Weight decay coefficient	5×10^{-4}

3.4. Model Comparison Experiment

To validate the effectiveness of the proposed TriRHC-YOLO model, comprehensive comparative experiments were conducted. First, considering that the proposed method is developed based on the YOLOv8n framework, YOLOv8n was selected as the baseline model to evaluate the effectiveness of the proposed improvement strategies. Second, four recently developed YOLO-series models, including YOLOv10n, YOLO11n, YOLO12, and YOLO26, were selected to assess the competitiveness of the proposed method against

advanced one-stage detectors. Subsequently, RT-DETR, a representative Transformer-based detection model, was introduced to compare the performance differences between convolution-based and attention-based detection architectures. Finally, six YOLO-based improved models, namely FFCA-YOLO [38], FBRT-YOLO [39], IF-YOLO [40], PKINet [41], LAM-YOLO [42], and LUD-YOLO [43], which are specifically designed for complex scenarios and small-object detection, were selected to further verify the effectiveness of the proposed method for UAV-based early-stage forest fire detection. All models were trained under the experimental settings described in Section 3.3. Their overall performance was quantitatively evaluated using three accuracy metrics—Recall, mAP@0.5, and mAP@0.5:0.95—together with model complexity and efficiency indicators, including the number of parameters (Params), computational complexity (GFLOPs), and inference speed (FPS). Specifically, Recall, mAP@0.5, mAP@0.5:0.95, Params, and GFLOPs were measured on the Ubuntu 20.04.5 LTS platform using a single NVIDIA vGPU (32 GB), whereas FPS was evaluated on the NVIDIA Jetson Orin NX embedded platform. The quantitative comparison results are presented in Table 3, and a visual comparison of model parameter sizes is shown in Figure 6.

Table 3. Results of Model Comparison Experiments.

Model	Recall	mAP@0.5	mAP@0.5:0.95	Para (M)	GFLOPs	FPS
YOLOv8n	0.711	0.794	0.577	3.167	8.912	209.783
YOLOv10n	0.691	0.782	0.561	2.784	8.729	174.955
YOLO11n	0.737	0.796	0.582	2.621	6.676	184.838
YOLO12	0.746	0.804	0.587	2.573	6.547	97.412
YOLO26	0.735	0.817	0.590	2.519	5.884	191.075
RT-DETR	0.686	0.737	0.489	32.810	108.321	27.336
FBRT-YOLO	0.749	0.801	0.565	0.885	6.882	162.243
FFCA-YOLO	0.726	0.812	0.580	8.496	72.974	34.574
IF-YOLO	0.744	0.813	0.592	10.588	26.453	41.982
LAM-YOLO	0.733	0.801	0.586	11.392	44.517	32.429
LUD-YOLO	0.750	0.806	0.587	3.914	18.565	28.710
PKINet	0.693	0.764	0.542	0.737	3.948	59.717
TriRHC-YOLO	0.769	0.848	0.608	3.181	9.732	168.251

Note: Bold values indicate the optimal results for each evaluation metric among the compared methods.

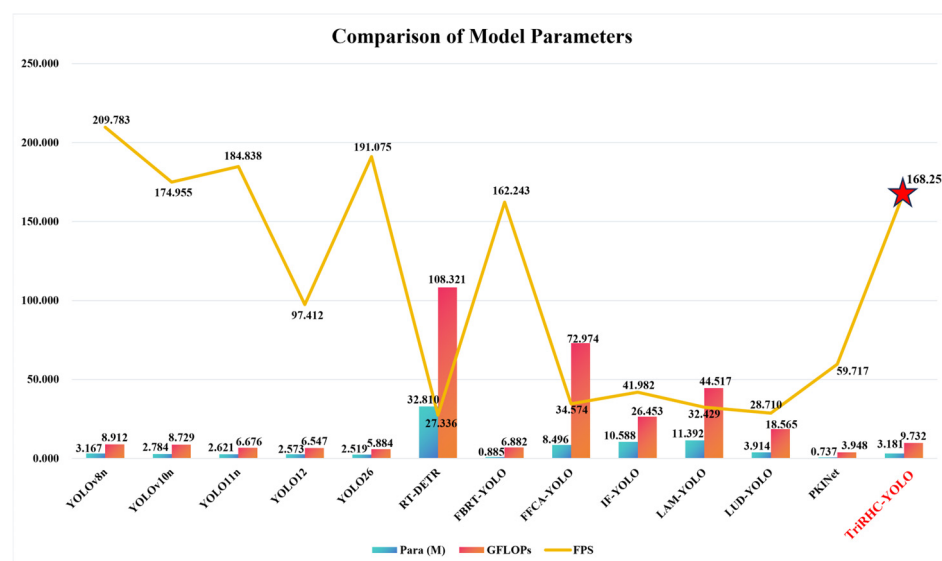


Figure 6. Visual Comparison of Model Parameter Features.

As shown in Table 3, the proposed TriRHC-YOLO achieves the best results in terms of Recall, mAP@0.5, and mAP@0.5:0.95. Specifically, the Recall reaches 0.769, representing an improvement of 8.16% over the baseline model YOLOv8n and 2.53% over the second-best model LUD-YOLO. This indicates that TriRHC-YOLO exhibits a stronger capability for capturing weak early flame targets under complex forest backgrounds and effectively reduces missed detections. In terms of target localization accuracy, TriRHC-YOLO achieves an mAP@0.5 of 0.848, which is 6.80% higher than that of the baseline model and 3.79% higher than that of the second-best model YOLO26. For the more stringent mAP@0.5:0.95 metric, TriRHC-YOLO achieves 0.608, which also outperforms all comparison models. These results demonstrate that the proposed model maintains stable localization performance under different confidence thresholds.

From the perspective of model architecture characteristics, although RT-DETR possesses strong global modeling capability, its parameter count reaches 32.810 M, with 108.321 GFLOPs and an FPS rate of only 27.336, making it difficult to satisfy the inference speed requirements of real-time UAV inspection scenarios. In terms of lightweight design and real-time performance, some models achieve higher inference speeds; however, their overall detection performance remains inferior to that of the proposed model. For example, YOLOv8n, YOLOv10n, YOLO11n, and YOLO26 all achieve higher FPS rates than the proposed model, yet their mAP@0.5:0.95 values are all lower than the 0.608 achieved by TriRHC-YOLO. In contrast, the proposed model achieves an FPS rate of 168.251. Although it is lower than that of YOLOv8n, YOLOv10n, YOLO11n, and YOLO26, it remains at a relatively high level and is sufficient to meet the engineering requirements of real-time detection on UAV onboard platforms.

To further visually demonstrate the detection performance of the proposed model in complex forest fire scenarios, visual comparative analyses were conducted between the proposed model and four representative models, namely PKINet, RT-DETR, IF-YOLO, and YOLO26, under different typical scenarios. The comparison results are shown in Figure 7.

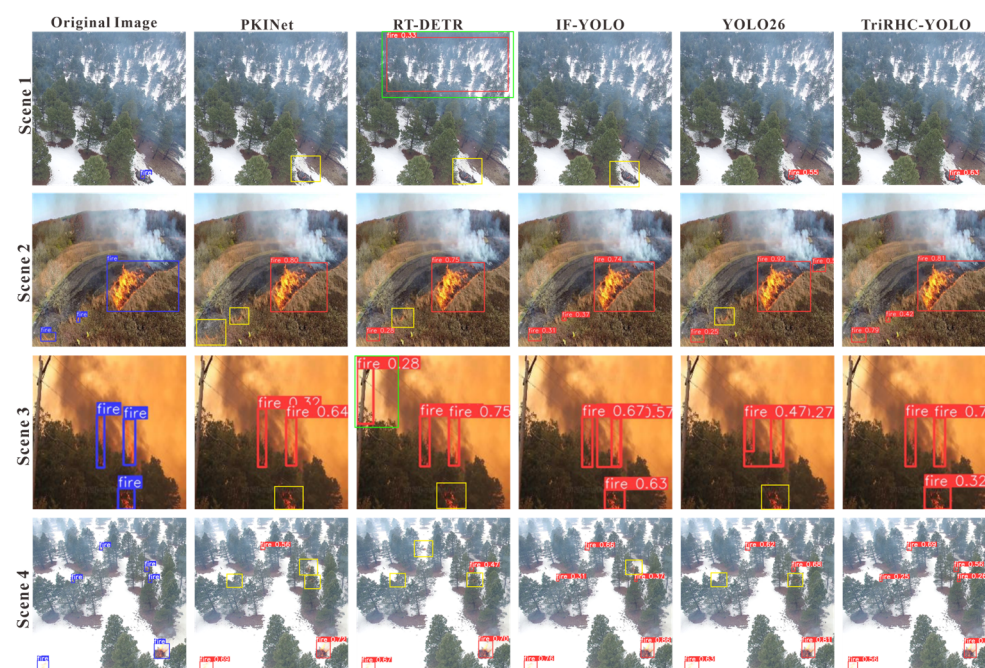


Figure 7. Visualization of Comparison Experiment Results. (Blue boxes indicate ground truth targets; Red boxes indicate prediction bounding boxes generated by each detection model; Yellow boxes indicate areas missed by the system; green boxes indicate areas incorrectly identified by the system).

As shown in Scene 1 of Figure 7, due to the high UAV imaging altitude, the fire-spot targets are small in scale and occupy only a limited proportion of image pixels. It can be observed that PKINet, IF-YOLO, and other models all exhibit missed detections, while RT-DETR produces relatively obvious false detections in the background. In contrast, TriRHC-YOLO is able to detect small-scale fire spots more completely and achieves higher detection confidence than YOLO26. This indicates that the proposed model can effectively preserve the shallow detailed features of tiny fire spots and possesses stronger feature-capturing capability for early weak fire spots under high-altitude viewing conditions.

As shown in Scene 2 of Figure 7, the textures of farmland bare soil and withered vegetation regions are highly complex, and the color of weak fire spots is highly similar to that of the background regions. PKINet, RT-DETR, and YOLO26 exhibit unstable recognition performance for small fire spots located in edge regions, resulting in missed detections to varying degrees. By comparison, TriRHC-YOLO can achieve stable detection of both the main flame body and weak fire spots located at the edges, demonstrating that the proposed method possesses strong feature discrimination capability and background suppression capability under complex backgrounds.

As shown in Scene 3 of Figure 7, the flame boundaries are blurred, the smoke transition regions are evident, and the continuity of the target contours is relatively poor. RT-DETR exhibits obvious false detections and missed detections in the background, whereas PKINet, IF-YOLO, and YOLO26 show incomplete detection of local weak flame regions and insufficient extraction of fine-grained flame contours. In comparison, TriRHC-YOLO is able to fully model fine-grained flame textures and high-level semantic features, thereby accurately extracting blurred flame boundaries. These results indicate that the proposed collaborative modeling strategy of directional, texture, and semantic features can enhance the model's representation capability for targets with blurred boundaries and weak textures.

As shown in Scene 4 of Figure 7, under a high-altitude UAV viewpoint, multiple weak-fire-spot targets are partially occluded by canopy vegetation and dispersed within forest backgrounds characterized by complex structures and diverse textures. When faced with such complex backgrounds, all four comparison models exhibit missed detections to varying degrees. By contrast, TriRHC-YOLO demonstrates outstanding detection capability for tiny fire spots under vegetation occlusion, effectively detecting a greater number of small-scale fire-spot targets while maintaining favorable localization accuracy. This reflects the stronger perception and representation capability of the proposed model for weak target features in complex scenarios.

To further evaluate the statistical stability and reliability of model performance, repeated experiments with different random seeds were conducted. Under identical dataset splits, training strategies, and experimental environments, YOLOv8n and TriRHC-YOLO were independently trained five times. The mean value, standard deviation (Std), and 95% confidence interval (95% CI) of Recall, mAP@0.5, and mAP@0.5:0.95 were calculated to quantify model robustness. The results are presented in Table 4.

Table 4. Stability Results under Different Random Seeds.

Model		YOLOv8n			TriRHC-YOLO	
Metric	Recall	mAP@0.5	mAP@0.5:0.95	Recall	mAP@0.5	mAP@0.5:0.95
Mean ± Std	0.714 ± 0.009	0.792 ± 0.006	0.579 ± 0.006	0.766 ± 0.006	0.847 ± 0.003	0.610 ± 0.004
95%CI	[0.703, 0.725]	[0.785, 0.799]	[0.572, 0.586]	[0.759, 0.774]	[0.843, 0.851]	[0.605, 0.615]

As shown in Table 4, TriRHC-YOLO consistently achieves higher performance than YOLOv8n under different random seeds. Compared with YOLOv8n, TriRHC-YOLO

obtains higher mean values for Recall, mAP@0.5, and mAP@0.5:0.95, indicating that the proposed improvements provide stable performance gains rather than accidental improvements caused by training randomness. Moreover, TriRHC-YOLO maintains relatively small standard deviations and narrow 95% confidence intervals across all evaluation metrics. For example, the mAP@0.5 of TriRHC-YOLO reaches 0.847 ± 0.003 with a 95% CI of [0.843, 0.851], demonstrating the robustness and reliability of the proposed model.

The comprehensive analysis demonstrates that TriRHC-YOLO can identify early weak flame targets more accurately and effectively under complex conditions, such as vegetation occlusion, blurred boundaries, and background interference, in UAV-based real-time forest fire detection tasks. Furthermore, it achieves a favorable balance among detection accuracy, model complexity, and real-time performance, making it more suitable for deployment in UAV-based real-time forest fire small-target detection scenarios.

3.5. Model Ablation Experiments

To verify the impact of the proposed RepVGG, HFA, and C2f-CStar modules on model performance, YOLOv8n was adopted as the baseline model. Ablation experiments were conducted by progressively introducing individual modules, dual-module combinations, and the three-module combination. The experimental results are presented in Table 5.

Table 5. Results of Module Ablation Experiments.

No.	Baseline	RepVGG	HFA	C2f-CStar	Recall	mAP@0.5	mAP@0.5:0.95	Para (M)
1	✓				0.711	0.794	0.577	3.163
2	✓	✓			0.706	0.789	0.582	2.874
3	✓		✓		0.724	0.803	0.585	3.352
4	✓			✓	0.718	0.798	0.584	3.208
5	✓	✓	✓		0.710	0.786	0.581	3.395
6	✓	✓		✓	0.717	0.789	0.575	2.923
7	✓		✓	✓	0.752	0.801	0.582	3.468
8	✓	✓	✓	✓	0.769	0.848	0.608	3.181

Note: ✓ Indicates that the corresponding module is adopted in the model configuration. Bold values indicate the optimal results for each evaluation metric within each group of ablation experiments.

From the results of the single-module ablation experiments, when the RepVGG module is introduced, the number of parameters decreases from 3.163 M to 2.874 M compared with the baseline model, and mAP@0.5:0.95 shows a slight improvement. However, Recall and mAP@0.5 exhibit slight declines, indicating that although this module reduces model complexity, relying solely on shallow local feature enhancement provides limited improvement in weak-target detection capability under complex forest backgrounds. After introducing the HFA module, Recall increases to 0.724, while mAP@0.5 and mAP@0.5:0.95 reach 0.803 and 0.585, respectively. These results indicate that the collaborative modeling of directional, texture, and semantic information can effectively enhance the model's discriminative capability for weak-target regions. Meanwhile, the number of parameters increases to 3.352 M, indicating that the attention mechanism introduces additional computational overhead. After introducing the C2f-CStar module, the overall performance of the model is improved compared with the baseline model, while the number of parameters only increases to 3.208 M. This verifies that the proposed module contributes to the detection of early small-scale fire spots under complex backgrounds.

From the results of the dual-module ablation experiments, certain differences can be observed among different module combinations. Specifically, the experimental results obtained by combining RepVGG with the other two modules are both lower than those of

the baseline model in terms of mAP@0.5, indicating that the collaborative effect between shallow feature enhancement and single high-level feature modeling is limited. In contrast, the combination of the HFA and C2f-CStar modules outperforms the baseline model across all evaluation metrics, demonstrating that C2f-CStar feature enhancement and the HFA mechanism exhibit strong complementarity.

After integrating all three modules, TriRHC-YOLO achieves Recall, mAP@0.5, and mAP@0.5:0.95 values of 0.769, 0.848, and 0.608, respectively, while maintaining a parameter count close to that of the baseline model. It achieves the best detection performance among all ablation experiments and realizes comprehensive improvements over the baseline model.

To further provide an intuitive analysis of the impact of each module on the model, Grad-CAM heatmap visualization analysis was conducted using the baseline YOLOv8n model (No. 1), the model with RepVGG introduced (No. 2), the model further incorporating HFA (No. 5), and the final TriRHC-YOLO model (No. 8). The results are shown in Figure 8.

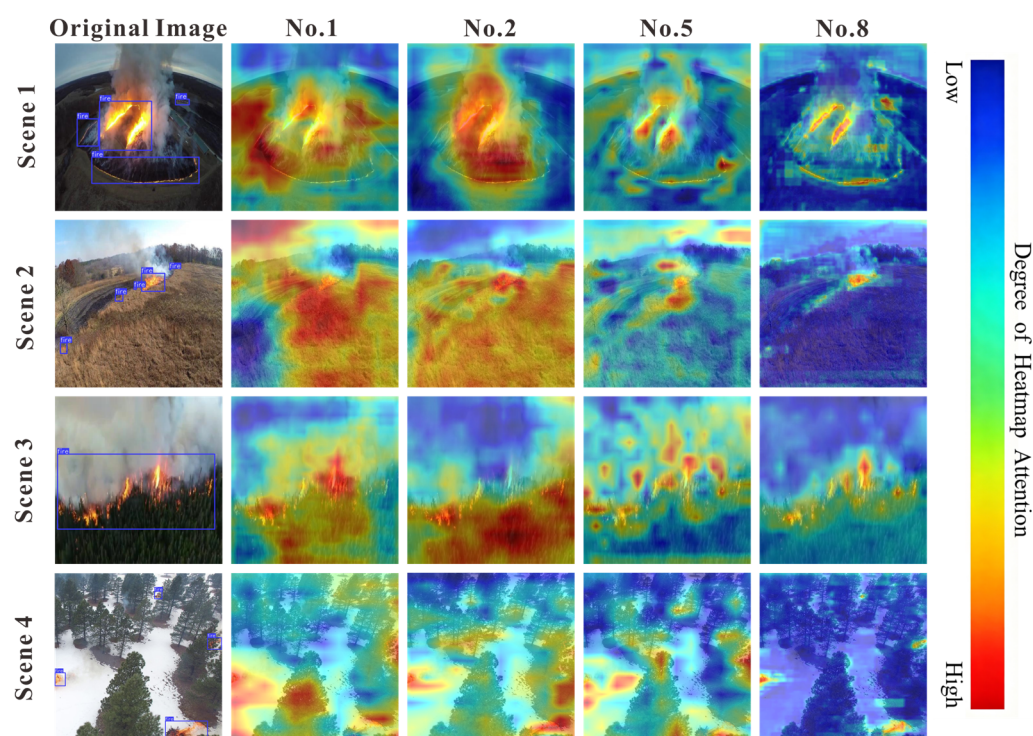


Figure 8. Visual Analysis of Module Ablation Heatmaps. (The blue box denotes the ground truth target label in the image).

In Scene 1, the fire line at the bottom exhibits weak texture characteristics and occupies a low proportion of image pixels. The baseline model shows weak responses to such shallow weak-feature targets, with attention distributed across multiple regions. After introducing RepVGG, the model pays greater attention to local high-temperature regions; however, it remains difficult to accurately focus on fine fire-line features. After incorporating the HFA module, the attention gradually shifts toward the main body of the fine fire line at the bottom. With the further introduction of C2f-CStar, the model can accurately identify the main fire-line region while eliminating interference from irrelevant areas.

In Scene 2, interference from backgrounds with similar color and texture characteristics, such as farmland and bare soil, causes the baseline model to exhibit strong responses over large background regions. After the successive introduction of RepVGG and HFA, the suppression of background interference is improved. The final model, however, mainly focuses on the actual fire-spot locations and can effectively avoid the negative effects caused by surface land-cover features.

In Scene 3, the flame boundaries are blurred, and the smoke transition regions are evident. The baseline model fails to effectively capture the characteristics of flames and smoke. The initial module combinations only slightly reduce the range of interference, exhibiting unstable responses at the flame–smoke transition regions and limited improvement in boundary discrimination capability. After introducing HFA, the model shows significantly enhanced attention to the flame body and smoke transition regions. Ultimately, the final model can clearly distinguish flames from smoke and demonstrates stronger perception capability for blurred boundaries and weak textures.

In Scene 4, under complex backgrounds involving occlusion, bare soil, and other interfering factors, the target responses of the baseline model exhibit obvious deviations and generate erroneous responses to certain background regions. As the modules are progressively introduced, the model's ability to suppress irrelevant backgrounds is improved; however, certain limitations in localizing concealed fire sources still remain. After integrating all modules, the model can accurately capture weak fire spots under occluded conditions while effectively suppressing interference from complex backgrounds, thereby exhibiting stronger adaptability to complex scenarios.

Overall, while controlling the number of parameters, the three modules collaboratively enable the model to focus more accurately on critical fire-spot regions in early forest fires, significantly improving its detection capability for early weak fire targets in complex scenarios.

4. Discussion

This study proposes the TriRHC-YOLO model based on the YOLOv8n architecture to address the challenges of small flame target scales, weak texture features, blurred boundaries, and complex backgrounds in UAV-based early forest fire detection.

From the perspective of the operating mechanisms of the proposed modules, RepVGG enhances shallow local detail features through structural reparameterization while reducing the number of model parameters. Particularly in Scene 1, which contains small-scale fire-line targets shown in heatmaps, the introduction of RepVGG enables the model to focus more on local high-temperature regions and flame-edge areas compared with YOLOv8n. However, according to the Recall and mAP@0.5 metrics, the performance improvement achieved by relying solely on the RepVGG module is limited. This indicates that weak fire-spot targets under complex forest backgrounds do not rely solely on shallow details and still require more advanced collaborative modeling of structural and semantic information. In Scene 3, a distinct transition region exists between flames and smoke, where the baseline model exhibits dispersed responses and regional confusion. The experimental results indicate that when HFA is introduced alone, Recall, mAP@0.5, and mAP@0.5:0.95 are significantly improved, demonstrating that HFA, through collaborative modeling of directional, texture, and semantic information across three branches, enables the model to simultaneously focus on continuous flame boundaries, local texture details, and high-level semantic information, thereby enhancing feature representation for blurred flame boundaries. However, when HFA is combined with RepVGG, mAP@0.5 and mAP@0.5:0.95 decrease, indicating that relying solely on shallow feature enhancement and attention mechanisms is insufficient to fully capture weak fire-spot targets in complex forest backgrounds, and further spatial feature reconstruction is still required. To address the characteristics of irregular flame diffusion and elongated fire lines, conventional convolutions often struggle to effectively capture global spatial correlations. In contrast, the C2f-CStar module achieves multi-scale contextual modeling through a cascaded star-shaped structure, compensating for the limitations of the network in spatial feature extraction. Meanwhile, the Star Multiplicative Gating mechanism further strengthens feature interaction across channels,

enabling the model to more effectively capture the features and local structural variations of weakly salient fire targets. The heatmap results in Scene 4 show that, under complex backgrounds and occlusion conditions, the introduction of C2f-CStar further enhances the model's focus on actual fire-spot regions while significantly reducing ineffective responses to background regions.

From the overall experimental results, TriRHC-YOLO achieves a favorable balance among detection accuracy, model complexity, and real-time performance. Compared with other comparison models, the proposed model does not rely on large-scale parameter expansion or high-complexity Transformer architectures. Instead, it achieves higher detection accuracy through targeted progressive structural optimization for the characteristics of early weak fires in UAV-based scenarios while maintaining only 3.181 M parameters. This makes it particularly suitable for airborne or edge-device-based UAV systems for early forest fire warning. The heatmap analysis further demonstrates that, with the progressive introduction of RepVGG, HFA, and C2f-CStar, the model's attention gradually shifts from background regions toward actual fire-spot regions, exhibiting more stable focus on weak fire targets in complex backgrounds involving smoke transitions and occlusions. This improvement not only increases the model's sensitivity to incipient fire targets but also suppresses spurious responses to irrelevant background features, thereby reducing the risk of false alarms in complex forest environments. This further indicates that, for UAV-based forest fire detection tasks, conducting structured improvements tailored to scene characteristics is more practically meaningful than simply increasing model scale.

Although TriRHC-YOLO improves the detection of weak early-stage fire targets in complex forest environments through feature enhancement and background suppression strategies, its practical performance may still be affected by challenging factors such as extreme weather conditions, varying illumination, and heavy smoke occlusion, which may increase the risk of false-positive and false-negative detections. In addition, the current model performs detection based on single-frame RGB images and does not fully exploit the temporal information inherent in fire evolution, potentially limiting its perception capability in dynamically changing fire scenarios.

Accordingly, several limitations remain to be addressed in future work. First, the current dataset mainly covers representative scenarios such as forests, farmland, and bare land; its ability to generalize to more challenging conditions, including extreme weather, complex illumination, and severe occlusion, requires further validation. Future studies will therefore focus on constructing a larger and more diverse forest fire dataset to further improve the robustness and generalization capability of the proposed model under complex environmental conditions. Second, the current model relies primarily on RGB imagery for fire detection. Future research could investigate RGB–thermal or other multimodal fusion frameworks by incorporating thermal infrared information to exploit temperature distribution characteristics, thereby further improving detection reliability for weak fires, low-contrast targets, and smoke-obscured environments. Finally, future work will incorporate temporal information from video sequences to fully exploit the dynamic characteristics of fire evolution, thereby enhancing the model's capability for continuous fire perception and improving the stability and early-warning performance of forest fire monitoring in complex scenarios.

5. Conclusions

To address the challenges of small fire-spot scales, insufficient feature representation of weak targets, blurred boundaries, and complex backgrounds in UAV-based forest fire inspection scenarios, this study proposes TriRHC-YOLO, an early forest fire detection model for complex environments based on YOLOv8n. The proposed model provides

progressive improvements targeting the feature representation deficiencies of fire-spot targets under different scenarios. First, a RepVGG module is introduced into the backbone network to enhance shallow edge feature extraction while reducing the number of model parameters, thereby achieving a lightweight design. Second, an HFA mechanism is designed to collaboratively improve the model's capability to recognize weak fire targets with blurred boundaries and complex backgrounds from three levels, namely direction, texture, and semantics. Finally, a C2f-CStar module is constructed. Through cascaded structural reconstruction and a Star Multiplicative Gating mechanism, the model's representation capability for small targets and weak fire targets is significantly enhanced.

In the comparative experiments, TriRHC-YOLO achieves the best Recall, mAP@0.5, and mAP@0.5:0.95 among all comparison models on the UAV-based early forest fire dataset. Specifically, compared with the baseline model YOLOv8n, Recall, mAP@0.5, and mAP@0.5:0.95 are improved by 8.16%, 6.80%, and 5.37%, respectively. In addition, the model contains only 3.181 M parameters, which is comparable to that of the baseline model, while achieving an FPS rate of 168.251. Therefore, the proposed model maintains a high inference speed while improving detection accuracy and satisfies the deployment requirements of UAV inspection scenarios. In the module ablation experiments, the model's response to fire-spot regions is progressively enhanced with the gradual introduction of the proposed modules, whereas the attention allocated to irrelevant background regions is significantly reduced. The overall model performance reaches its optimum when all three modules are integrated. The heatmap results further demonstrate that TriRHC-YOLO can more accurately focus on small fire spots, smoke transition regions, and weak fire spots under complex backgrounds, effectively reducing the risks of missed detections and false detections. These results further verify the collaborative effectiveness of the RepVGG, HFA, and C2f-CStar modules.

Overall, TriRHC-YOLO achieves a favorable balance between model complexity and real-time detection performance while substantially improving the detection capability for weak early-stage fire targets in complex forest environments. Through the complementary effects of the three feature enhancement modules, the proposed model effectively reduces both false-positive and false-negative detections under challenging environmental conditions, thereby enhancing the reliability of UAV-based intelligent forest fire inspection systems. In addition, it minimizes the need for secondary manual verification caused by false alarms and reduces the unnecessary consumption of firefighting personnel and material resources, ultimately providing more efficient information support for emergency responses.

Author Contributions: Conceptualization, B.S. and B.L.; methodology, B.S., B.L. and Z.Z.; experimentation, B.S., Q.W. and X.Z.; validation, B.S., B.L., Z.Z., Q.W., X.Z., Z.C., T.Y. and J.J.; investigation, B.L., Z.Z., T.Y. and J.J.; resources, Y.C.; data curation, Q.W., X.Z. and T.Y.; writing—original draft preparation, B.S. and B.L.; writing—review and editing, Y.C.; visualization, Z.Z., Z.C. and J.J.; supervision, Y.C.; project administration, Y.C.; funding acquisition, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guangxi Youth Talent Program and supported in part by the Guangxi Science and Technology Project (Grant Nos. Guike FN2600640051 and Guike AB25069093) and the Guangxi Key Laboratory of Ecological Spatiotemporal Big Data Perception Service Program (Grant No. GXKEY2026B08).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are from the publicly available FLAME dataset published by Shamsoshoara A. et al. (2021) [36] and the publicly available FlameVision dataset published by ANAM I J. (2023) [37]. They are available at: <https://ieee-dataport.org/open-access/flame-dataset-aerial-imagery-pile-burn-detection-using-drones-uavs> (accessed on 1 June 2026) and <https://www.kaggle.com/datasets/anamibnjafar0/flamevision> (accessed on 1 June 2026).

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

Abbreviations

The following abbreviations are used in this manuscript:

UAV	Unmanned Aerial Vehicle
TP	True Positive
FN	False Negative
AP	Average Precision
mAP	mean Average Precision
RepVGG	Reparameterized VGG
HFA	Hierarchical Feature Attention
C2f-CStar	Cross Stage Partial Feature Fusion with Cascade Star Block
DWConv	Depthwise Convolution
CNNs	Convolutional Neural Networks
GAP	Global Average Pooling
FC	Fully Connected
Para	Parameters
GFLOPS	Giga Floating-Point Operations Per Second
FPS	Frames Per Second

References

1. Yang, Z.; Shao, Y.; Wei, Y.; Li, J. Precision-boosted forest fire target detection via enhanced YOLOv8 model. *Appl. Sci.* **2024**, *14*, 2413. [\[CrossRef\]](#)
2. Zhao, J.; Yue, C.; Wang, J.; Hantson, S.; Wang, X.; He, B.; Li, G.; Wang, L.; Zhao, H.; Luyssaert, S. Forest fire size amplifies postfire land surface warming. *Nature* **2024**, *633*, 828–834. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Maeda, N.; Tonooka, H. Early stage forest fire detection from Himawari-8 AHI images using a modified MOD14 algorithm combined with machine learning. *Sensors* **2023**, *23*, 210. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Manoj, S.; Valliyammai, C. Drone network for early warning of forest fire and dynamic fire quenching plan generation. *EURASIP J. Wirel. Commun. Netw.* **2023**, *2023*, 112. [\[CrossRef\]](#)
5. Liu, X.; Chen, Z.; Jiang, W. YOLO-MC: An algorithm for early forest fire recognition based on drone image. *Open Geosci.* **2025**, *17*, 20250875. [\[CrossRef\]](#)
6. Chen, Y.; Li, J.; Sun, K.; Zhang, Y. A lightweight early forest fire and smoke detection method. *J. Supercomput.* **2024**, *80*, 9870–9893. [\[CrossRef\]](#)
7. Ali, A.W.; Kurnaz, S. Optimizing deep learning models for fire detection, classification, and segmentation using satellite images. *Fire* **2025**, *8*, 36. [\[CrossRef\]](#)
8. Khan, T.; Singh, K.; Bhati, B.S.; Ahmad, K.; Al-Rasheed, A.; Getahun, M.; Soufiene, B.O. Trust-driven approach to enhance early forest fire detection using machine learning. *Sci. Rep.* **2025**, *15*, 14480. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Luo, Z.; Xu, H.; Xing, Y.; Zhu, C.; Jiao, Z.; Cui, C. YOLO-UFS: A novel detection model for UAVs to detect early forest fires. *Forests* **2025**, *16*, 743. [\[CrossRef\]](#)
10. Yu, P.; Wei, W.; Li, J.; Du, Q.; Wang, F.; Zhang, L.; Li, H.; Yang, K.; Yang, X.; Zhang, N.; et al. Fire-PPYOLOE: An efficient forest fire detector for real-time wild forest fire monitoring. *J. Sens.* **2024**, *2024*, 2831905. [\[CrossRef\]](#)
11. Wu, H.; Li, H.; Shamsoshoara, A.; Razi, A.; Afghah, F. Transfer learning for wildfire identification in UAV imagery. In Proceedings of the 2020 54th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 18–20 March 2020; pp. 1–6.
12. Barmpoutis, P.; Papaioannou, P.; Dimitropoulos, K.; Grammalidis, N. A review on early forest fire detection systems using optical remote sensing. *Sensors* **2020**, *20*, 6442. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Pagar, P.B.; Shaikh, A.N. Real time based fire and smoke detection without sensor by image processing. *Int. J. Adv. Electr. Electron. Eng.* **2013**, *2*, 25–34.

14. Çelik, T.; Özkaramanlı, H.; Demirel, H. Fire and smoke detection without sensors: Image processing based approach. In Proceedings of the 2007 15th European Signal Processing Conference, Poznań, Poland, 3–7 September 2007; pp. 1794–1798.
15. Çelik, T.; Demirel, H.; Ozkaramanlı, H.; Uyguroglu, M. Fire detection using statistical color model in video sequences. *J. Vis. Commun. Image Represent.* **2007**, *18*, 176–185. [\[CrossRef\]](#)
16. Çelik, T.; Demirel, H. Fire detection in video sequences using a generic color model. *Fire Saf. J.* **2009**, *44*, 147–158. [\[CrossRef\]](#)
17. Töreyn, B.U.; Dedeoğlu, Y.; Cetin, A.E. Wavelet based real-time smoke detection in video. In Proceedings of the 2005 13th European Signal Processing Conference, Antalya, Turkey, 4–8 September 2005; pp. 1–4.
18. Chen, T.H.; Wu, P.H.; Chiou, Y.C. An early fire-detection method based on image processing. In Proceedings of the 2004 International Conference on Image Processing (ICIP 2004), Singapore, 24–27 October 2004; Volume 3, pp. 1707–1710.
19. Nguyen-Ti, T.; Nguyen-Phuc, T.; Do-Hong, T. Fire detection based on video processing method. In Proceedings of the 2013 International Conference on Advanced Technologies for Communications (ATC 2013), Ho Chi Minh City, Vietnam, 16–18 October 2013; pp. 106–110.
20. Qiu, T.; Yan, Y.; Lu, G. An autoadaptive edge-detection algorithm for flame and fire image processing. *IEEE Trans. Instrum. Meas.* **2011**, *61*, 1486–1493. [\[CrossRef\]](#)
21. Huang, J.; Zhou, J.; Yang, H.; Liu, Y.; Liu, H. A small-target forest fire smoke detection model based on deformable transformer for end-to-end object detection. *Forests* **2023**, *14*, 162. [\[CrossRef\]](#)
22. Zheng, Y.; Zhang, G.; Tan, S.; Yang, Z.; Wen, D.; Xiao, H. A forest fire smoke detection model combining convolutional neural network and vision transformer. *Front. For. Glob. Chang.* **2023**, *6*, 1136969. [\[CrossRef\]](#)
23. Rao, G.N.; Rao, P.J.; Duvvuru, R.; Beulah, K.; Lydia, E.L.; Rathnala, P.; Balakrishna, B.; Motru, V.R. Neural fuzzy system design in forest fire detection. *Microsyst. Technol.* **2024**, *30*, 455–467. [\[CrossRef\]](#)
24. Cheknane, M.; Bendouma, T.; Boudouh, S.S. Advancing fire detection: Two-stage deep learning with hybrid feature extraction using faster R-CNN approach. *Signal Image Video Process.* **2024**, *18*, 5503–5510. [\[CrossRef\]](#)
25. He, L.; Zhou, Y.; Liu, L.; Zhang, Y.; Ma, J. Research and application of deep learning object detection methods for forest fire smoke recognition. *Sci. Rep.* **2025**, *15*, 16328. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Frizzi, S.; Kaabi, R.; Bouchouicha, M.; Ginoux, J.; Moreau, E.; Fnaiech, F. Convolutional Neural Network for Video Fire and Smoke Detection. In Proceedings of the IECON 2016–42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 23–26 October 2016. [\[CrossRef\]](#)
27. Li, C.; Du, Y.; Zhang, X.; Wu, P. YOLOGX: An improved forest fire detection algorithm based on YOLOv8. *Front. Environ. Sci.* **2025**, *12*, 1486212. [\[CrossRef\]](#)
28. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10778–10787. [\[CrossRef\]](#)
29. Pang, Y.; Wu, Y.; Yuan, Y. FuF-Det: An early forest fire detection method under fog. *Remote Sens.* **2023**, *15*, 5435. [\[CrossRef\]](#)
30. Yar, H.; Hussain, T.; Agarwal, M.; Khan, Z.A.; Gupta, S.K.; Baik, S.W. Optimized dual fire attention network and medium-scale fire classification benchmark. *IEEE Trans. Image Process.* **2022**, *31*, 6331–6343. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Yang, W.; Yang, Z.; Wu, M.; Zhang, G.; Zhu, Y.; Sun, Y. SIMCB-Yolo: An efficient multi-scale network for detecting forest fire smoke. *Forests* **2024**, *15*, 1137. [\[CrossRef\]](#)
32. Dilshad, N.; Pareez, S.; Khan, S.U.; Khan, M.; Mahmood, T.; Alghamdi, N.S.; Song, J. YOLOFlame: Benchmarking vehicle fire and smoke detection in urban environments. *Alex. Eng. J.* **2026**, *141*, 109–118. [\[CrossRef\]](#)
33. Gao, L.; Chen, G. YOLO11-RLN: An aerial UAV algorithm for forest fire detection. *Ann. N. Y. Acad. Sci.* **2025**, *1551*, 312–324. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Shamta, I.; Demir, B.E. Development of a deep learning-based surveillance system for forest fire detection and monitoring using UAV. *PLoS ONE* **2024**, *19*, e0299058. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Ding, X.; Zhang, X.; Ma, N.; Han, J.; Ding, G.; Sun, J. RepVGG: Making VGG-style convnets great again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 13733–13742.
36. Shamsoshoara, A.; Afghah, F.; Razi, A.; Zheng, L.; Fulé, P.Z.; Blasch, E. Aerial imagery pile burn detection using deep learning: The FLAME dataset. *Comput. Netw.* **2021**, *193*, 108001. [\[CrossRef\]](#)
37. Anam, I.J. FlameVision. Available online: <https://www.kaggle.com/datasets/anamibnjafar0/flamevision> (accessed on 1 June 2026).
38. Zhang, Y.; Ye, M.; Zhu, G.; Liu, Y.; Guo, P.; Yan, J. FFCA-YOLO for small object detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5611215. [\[CrossRef\]](#)
39. Xiao, Y.; Xu, T.; Xin, Y.; Li, J. FBRT-YOLO: Faster and better for real-time aerial image detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025; Volume 39, pp. 8673–8681. [\[CrossRef\]](#)
40. Zhang, J.; Zhang, Y.; Shi, Z.; Zhang, Y.; Gao, R. Unmanned aerial vehicle object detection based on information-preserving and fine-grained feature aggregation. *Remote Sens.* **2024**, *16*, 2590. [\[CrossRef\]](#)

41. Cai, X.; Lai, Q.; Wang, Y.; Wang, W.; Sun, Z.; Yao, Y. Poly kernel inception network for remote sensing detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2024; pp. 27706–27716.
42. Zheng, Y.; Jing, Y.; Zhao, J.; Cui, G. LAM-YOLO: Drones-based small object detection on lighting-occlusion attention mechanism YOLO. *Comput. Vis. Image Underst.* **2025**, *261*, 104489. [[CrossRef](#)]
43. Fan, Q.; Li, Y.; Deveci, M.; Zhong, K.; Kadry, S. LUD-YOLO: A novel lightweight object detection network for unmanned aerial vehicle. *Inf. Sci.* **2025**, *686*, 121366. [[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.