



Article

FuF-Det: An Early Forest Fire Detection Method under Fog

Yaxuan Pang , Yiquan Wu * and Yubin Yuan

School of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; hins_pang@nuaa.edu.cn (Y.P.); harley_yuan@nuaa.edu.cn (Y.Y.)

* Correspondence: imagestrong@nuaa.edu.cn; Tel.: +86-137-7666-7415

Abstract: In recent years, frequent forest fires have seriously threatened the earth's ecosystem and people's lives and safety. With the development of machine vision and unmanned aerial vehicle (UAVs) technology, UAV monitoring combined with machine vision has become an important development trend in forest fire monitoring. In the early stages, fire shows the characteristics of a small fire target and obvious smoke. However, the presence of fog interference in the forest will reduce the accuracy of fire point location and smoke identification. Therefore, an anchor-free target detection algorithm called FuF-Det based on an encoder–decoder structure is proposed to accurately detect early fire points obscured by fog. The residual efficient channel attention block (RECAB) is designed as a decoder unit to improve the problem of the loss of fire point characteristics under fog caused by upsampling. Moreover, the attention-based adaptive fusion residual module (AAFRM) is used to self-enhance the encoder features, so that the features retain more fire point location information. Finally, coordinate attention (CA) is introduced to the detection head to make the image features correspond to the position information, and improve the accuracy of the algorithm to locate the fire point. The experimental results show that compared with eight mainstream target detection algorithms, FuF-Det has higher average precision and recall as an early forest fire detection method in fog and provides a new solution for the application of machine vision to early forest fire detection.

Keywords: early forest fire detection; forest fog; anchor-free; adaptive feature enhancement; deep learning



Citation: Pang, Y.; Wu, Y.; Yuan, Y. FuF-Det: An Early Forest Fire Detection Method under Fog. *Remote Sens.* **2023**, *15*, 5435. <https://doi.org/10.3390/rs15235435>

Academic Editor: Ioannis Gitas

Received: 8 October 2023

Revised: 6 November 2023

Accepted: 10 November 2023

Published: 21 November 2023



Copyright: © 2023 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) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Forests, known as the “lungs of the Earth”, possess significant ecological and economic value. Indeed, habitats for 80% of the Earth's organisms are provided by forests. The abundant biodiversity in forests is crucial for maintaining ecosystem stability. Moreover, the green vegetation in forests purifies harmful gases such as sulfur dioxide and chlorine, improving atmospheric conditions. Additionally, plant photosynthesis balances the oxygen and carbon dioxide levels in the air, contributing to mitigating global warming [1]. Furthermore, forests offer abundant resources such as timber, medicinal plants, and food materials, vital to human economic and social development. However, in recent years, forests have been increasingly affected by abnormal weather conditions and human activities, leading to frequent forest fires. In 2009, bushfires in Victoria, Australia, killed 173 people. In 2018, wildfires in California were responsible for the deaths of 85 people [2]. For rural areas located in the forest, a sudden forest fire can destroy villagers' homes, agricultural assets and other infrastructure. At the same time, it also has an irreversible impact on people's health and lives. The 2003 bushfire in Australia's Wulgulmerang region notably destroyed the sparsely populated farming region [3]. In 2019, the Amazon rainforest fires consumed millions of hectares of land, releasing significant carbon dioxide and carbon monoxide, severely impacting global air quality. According to statistics, from 2001 to 2022, the total reduction in tree cover globally due to fires amounted to 126 million hectares [4]. These figures serve as a wake-up call for taking effective measures to prevent and promptly extinguish forest fires.

Forest fires are extremely destructive. All countries in the world try their best to prevent or monitor forest fires as soon as possible and stop their spread. In the fire prevention phase, prescribed fire is one of the tools used to manage fire. Prescribed fire refers to the deliberate ignition of potential fuels in the forest under the conditions of maximum temperature, relative humidity and wind speed at the threshold required for fire to spread [5] in order to reduce the fuel density in the forest and prevent the occurrence of destructive forest fires [6]. For forest fire monitoring, common methods can be categorized into four types: manual patrols, sensor-based monitoring [7–9], fire satellite monitoring [10–14], and UAV monitoring [15–18]. Manual patrols are inefficient, have limited coverage, and expose humans to potential risks in hazardous environments [19]. Sensors detect the presence of fires by monitoring indicators such as smoke, humidity, and temperature. However, relying solely on these indicators can lead to false alarms due to the complexity of forest environments. Moreover, the installation cost and limited coverage of sensors make them unsuitable for forest fire monitoring [20,21]. Fire satellites predict fires by monitoring changes in surface temperature. They can cover areas inaccessible to humans and have a wide coverage range. However, satellite monitoring faces challenges in achieving ideal spatial and temporal resolutions simultaneously, which makes this method unsuitable for early forest fire detection, which requires real-time monitoring and high detection rates for small fire points [20]. In contrast, UAV monitoring has attracted more researchers' attention due to the small size of the equipment involved, its light weight, and high flexibility. Although the current load supported by UAVs is not high, resulting in a limited detection area and duration, UAVs equipped with camera sensors can provide higher resolution images and real-time monitoring capabilities, so the application of UAV remote sensing technology to forest fire detection is the future development trend [22].

With the advancement of computer vision, there has been an increasing number of cases where UAVs combined with computer vision techniques are applied to forest fire detection. Early forest fires are characterized by small fire points and distinct smoke features. Therefore, relevant research can be divided into two directions: 1. Smoke detection in early forest fires. 2. Detection of small fire points in early forest fires.

1.1. Smoke Detection

Computer vision utilizes feature extraction to recognize the targets of interest in detection. Initially, researchers employed traditional methods to extract smoke features, such as local binary patterns, histograms of oriented gradients, wavelet analysis, principal component analysis, etc. [23–26]. However, due to the complexity of forest environments, features extracted through traditional methods have poor robustness and are easily affected by complex backgrounds, resulting in subpar detection performance. With the development of deep learning techniques, features extracted using deep learning methods exhibit more robust representation capabilities for the target, enabling better detection of smoke in complex forest environments [27,28]. Wu et al. [29] utilized pulse-coupled neural networks to extract texture features of smoke and combined them with support vector machines for smoke classification. Lu [30] designed DarkCNN, which extracts images' dark channel features, improving the performance of forest fire smoke detection. Experimental results have shown that the dark channel features of smoke enhance the detection performance of various convolutional neural network (CNN) models.

In practical scenarios, aerial images captured by UAVs inevitably contain clouds, fog, or other smoky targets. Moreover, smoke exhibits dynamic changes, with diffusion over time leading to problems such as semi-transparency and lack of concentration. These factors further complicate the detection process. He et al. [31] combined spatial and channel attention mechanisms with VGG16 to allow the network to focus on smoke characteristics, thus avoiding interference from complex forest backgrounds. They also employed feature fusion techniques to preserve global image features and improve missed detection for small smoke targets. Li et al. [32] improved early smoke detection by adding the convolutional block attention module (CBAM) to the network used for feature extraction. Addressing the

issue of irregular smoke diffusion and unclear smoke features caused by complex forest environments, Hu et al. [33] designed a joint weight allocation strategy based on horizontal and vertical directions using an attention mechanism called VAM to extract smoke texture features. Similarly, Zhang et al. [34] proposed MMFNet based on CNN. They first designed a mixed attention to highlight the effects of smoke's horizontal and vertical texture features. They then constructed a multi-scale convergence coordinated pyramid network to fuse features extracted by CNN at different scales. Finally, by combining prediction heads, they achieved multi-scale smoke detection.

With the emergence of various target detection algorithms, You Only Look Once (YOLO), a fast one-stage detection algorithm, has gained popularity among researchers. Studies have shown that compared to two-stage target detection algorithms, SSD and YOLO series one-stage detection algorithms are more suitable for detecting smoke in forest fires [35,36]. Zhan et al. [16] proposed ARGNet based on the PP-YOLO algorithm to address the issues of high transparency and unclear edges after smoke diffusion. They first used SoftPool in the feature extraction stage to avoid the loss of transparent smoke features. In the feature pyramid network (FPN) structure of the neck, they employed deconvolution and dilated convolution for global feature fusion, enhancing the algorithm's ability to detect small smoke targets. Finally, they introduced the global optimal non-maximum suppression method to replace the matrix NMS in the original algorithm, enabling multi-target detection in a single image. Li et al. [37] proposed ALFRNet, an adaptive linear feature reuse network for rapid smoke detection in forest fires. They used adaptive depth convolution and residual connection structures to construct a bilinear feature reuse module, addressing the issue of feature loss in transparent smoke during downsampling. Additionally, they incorporated a hybrid attention-guided module using CBAM residual connections to highlight the features of transparent smoke. The introduction of these techniques improved the issue of missed detection in the YOLOv3 algorithm during fire and smoke detection. However, these algorithms struggle to handle scenarios where fire smoke coexists with atmospheric phenomena such as clouds and fog. Li et al. [38] proposed a recursive BiFPN attention algorithm called RBiFPN for secondary fusion of fused features. This would help the algorithm focus on smoke features. They integrated RBiFPN as the feature fusion module in YOLOv5. Additionally, they employed swin-TPH as the detection head, utilizing its hierarchical structure to enhance the algorithm's ability to detect small smoke targets. Qian et al. [39] introduced the omni-dimensional dynamic convolution and bottleneck transformer structures into the YOLOv5 backbone. This allowed the algorithm to pay more attention to global features in the feature extraction process. They also incorporated the SimAM attention mechanism before the prediction head to enhance the algorithm's ability to detect small targets. Experimental results demonstrate that this algorithm can effectively distinguish between clouds and smoke.

Smoke and fog can be distinguished based on the characteristics of satellite remote sensing in the thermal infrared band when fire smoke is obscured by fog. However, since it is difficult to acquire and process high-resolution remote sensing images and the temporal resolution of remote sensing images is not sufficient to meet the characteristics of instantaneous fire occurrence, the method proposed in this paper makes up for the shortcomings of remote sensing methods [20].

1.2. Fire Point Detection

Due to the small size of early forest fire points and the fact that UAVs operate at certain flying heights, the difficulty of early forest fire detection is further increased [15]. Forest fire detection based on infrared imagery can effectively utilize the temperature differences between fire points and the surrounding environment, providing high accuracy and real-time capabilities [40,41]. However, due to the high cost of infrared imaging devices, there are better options for researchers. In early forest fire detection methods, researchers often extract flame regions in different color spaces, such as YCbCr, HSV, HSL, HWB, and Lab, combined with machine learning classifiers to achieve fire point detection [42–45].

However, obstacles such as leaves and smoke can weaken the color characteristics of flames, and strong sunlight can also interfere with the color space features of flames.

With the development of deep learning technology, researchers are no longer limited to features such as the color of flames. Instead, they use CNN to extract deep fire features, reducing the false positive detection rate [46–48]. Various target detection algorithms have gradually matured and have been widely applied in forest fire detection tasks in recent years. For two-stage target detection algorithms, Zhang et al. [17] proposed the MS-FRCNN algorithm to address the early-stage detection of small fire points. Based on Faster R-CNN, it incorporates the FPN structure to fuse multi-scale information and introduces an attention mechanism into the region proposal network. By employing spatial parallel attention mechanisms, the algorithm focuses on the regions where small targets are located, avoiding the missed detection of small fire points. However, it turns out that two-stage target detection algorithms have slower detection speeds. In the context of forest fire detection tasks, detecting forest fires more quickly can provide better opportunities for disaster relief personnel to extinguish fires. Compared with two-stage target detection algorithms, one-stage target detection algorithms do not involve candidate region generation steps. They directly detect targets in input images, making them more suitable for forest fire detection tasks. As a representative of one-stage detection algorithms, combining YOLO and UAVs enables high frame rate detection. However, direct use of YOLO is prone to missed detections [49]. To address this, most studies have enhanced the attention and feature fusion modules of the YOLO algorithm to improve its focus on small fire points and reduce the missed detection rate [50–54]. To address the issue of false positives, Zheng et al. [18] proposed a preliminary detection-enhancement secondary detection approach to reduce the false positive rate of YOLOv4 in forest fire detection. However, the YOLO target detection algorithm is sensitive to target size. When the size of the detected target does not match that of the training target, the algorithm struggles to obtain accurate detection boxes. To enhance the accuracy of fire point localization, some studies have adopted encoder–decoder structures with greater flexibility for fire point detection. Lu et al. [55] designed the MTL-FFDET algorithm based on an encoding–decoding structure for multi-task learning in forest fire detection, enabling simultaneous fire classification, segmentation, and detection tasks. The above forest fire points detection algorithms are all anchor-based, and few algorithms take into account the problem of smoke and fog occlusion. Huang et al. [56] combined the defogging algorithm and YOLOX to propose an algorithm, GXLD, that can be used to detect forest fires in fog scenes. The test results show the feasibility of forest fire detection on foggy days, but GXLD still has some limitations in severe fog scenes.

Based on the above, most deep learning forest fire detection algorithms are anchor-based. However, in practical scenarios, predefined anchor boxes struggle to capture the information of small fire points in the early stages. In addition, many forest fire detection algorithms determine the occurrence of forest fires by detecting smoke or flames. However, fog is common in forests due to the transpiration of numerous plants, as shown in Figure 1. Although the humidity is higher on foggy days, which means the probability of a fire occurring naturally is lower, the existence of ground fires and human factors (such as outdoor smoking, heating and cooking, burning incense, etc.) can still lead to forest fires. When a significant amount of fog is present, the smoke generated by the fires becomes difficult to capture by UAVs, and the fog obscures the features of the flames. Addressing the above issues, FuF-Det, an anchor-free detection algorithm based on an encoder–decoder structure, is proposed. Our specific contributions are as follows:

1. The FuF-Det algorithm was proposed to enhance the detection accuracy of early forest fires in foggy scenes.
2. To preserve the positional features of early fire points during the downsampling process of the encoder, AAFRM is designed as a feature fusion structure between the encoder and decoder.
3. To address the issue of losing fine-grained fire point details in the presence of fog during upsampling, RECAB is constructed as the decoder unit.

4. To enhance the accuracy of early forest fire localization, CA is introduced into the anchor-free detection head, resulting in the CA-Head.



Figure 1. Examples of forest fog: (a–c) mild, moderate, and severe fog scenes, respectively [57].

2. Methods

2.1. FuF-Det Algorithm

Early fire detection in foggy scenes faces two challenges. First, there is a heavy fog that hides the most noticeable color of the flames. Second, early fires tend to have a smaller scale, and UAVs capturing images from certain heights result in smaller fire targets in the images. To address the issue of early forest fire detection in foggy scenes, an anchor-free detection algorithm based on an encoder–decoder architecture named FuF-Det is proposed. The algorithm consists of four components: encoder, decoder, feature fusion structure, and detection head. The encoder efficiently extracts semantic information from images by using ResNet50 [58] with residual connections. The decoder, composed of RECAP modules with different input–output sizes, avoids losing fire point details during upsampling. The feature fusion structure integrates the encoder features self-enhanced by AAFRM with the decoder features, enhancing the fire point localization information. AAFRM enhances the position information of fire points in both channel and spatial dimensions. Finally, the output of the decoder is enhanced by the CA-Head, which leverages the correlation between image features and their corresponding positional information to improve the detection head’s ability to perceive the precise locations of small targets. Figure 2 shows the structure of the FuF-Det algorithm.

2.2. Encoder

Within the encoder–decoder architecture, the encoder plays a crucial role in transforming the input image into a deeper representation of features through downsampling. This process compresses and abstracts the original image, capturing relevant information while reducing redundancy. ResNet50 is a deep convolutional neural network known for its residual connections, facilitating rapid information propagation and skipping, and promoting feature transfer and reuse. This architecture enables the network to capture features and semantic information at different levels, making it well suited for extracting fire point characteristics in fog obscuration. In addition, the residual connections allow deeper network structures while reducing the number of parameters and computations. This results in a lightweight network that can be easily embedded in UAVs for practical detection in real-world scenarios. The structure and composition of the encoder are illustrated in Table 1.

ResNet50 consists mainly of two modules: Conv Block and Identity Block. Their structures are shown in Figure 3. Both modules consist of three convolutional layers with kernel sizes of 1×1 , 3×3 , and 1×1 , along with BatchNorm (BN) layers and ReLU activation functions. The main difference lies in the residual branch: Conv Block includes an additional convolutional layer with a 1×1 kernel size and a BN layer to modify the feature dimensions. At the same time, the Identity Block maintains the same input and

output dimensions, deepening the network. In FuF-Det, the input image size for the encoder is set to $512 \times 512 \times 3$. In the C1 stage of the network, the input image undergoes a 7×7 convolutional layer with a stride of 2, followed by a BN layer and a 3×3 max pooling layer with a stride of 2 to achieve downsampling and a reduced image size. Stages C2–C5 consist of a Conv Block and several Identity Blocks, extracting deeper image features. Finally, the encoder abstracts the input image with dimensions of $512 \times 512 \times 3$ into a feature map with dimensions of $16 \times 16 \times 2048$.

Table 1. Encoder structure.

Stage	Module	Module Numbers	Output
C1	Conv2d	1	$256 \times 256 \times 64$
	BatchNorm	1	-
	Maxpool	1	$128 \times 128 \times 64$
C2	Conv Block	1	$128 \times 128 \times 256$
	Identity Block	2	
C3	Conv Block	1	$64 \times 64 \times 512$
	Identity Block	3	
	Identity Block	3	
C4	Conv Block	1	$32 \times 32 \times 1024$
	Identity Block	5	
C5	Conv Block	1	$16 \times 16 \times 2048$
	Identity Block	2	

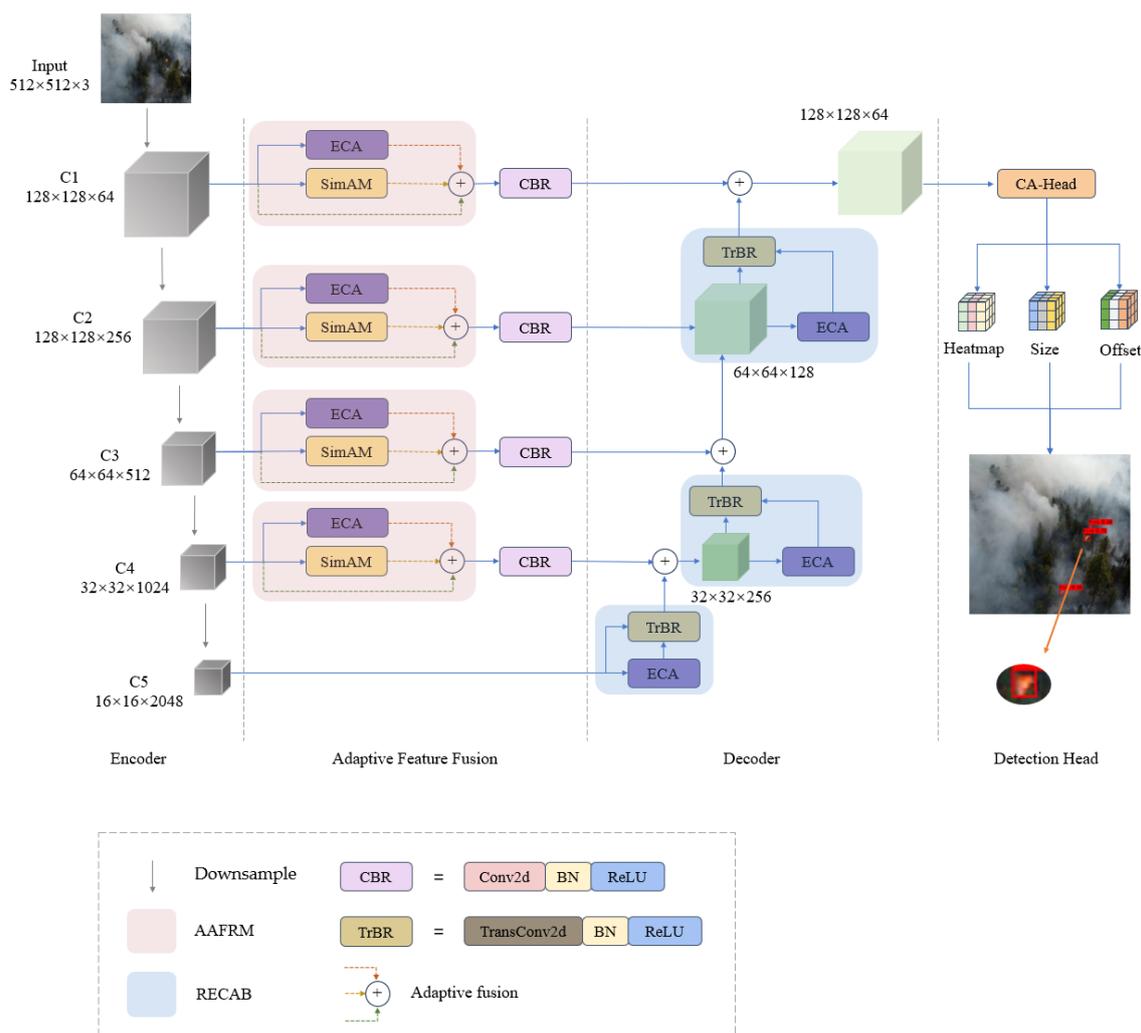


Figure 2. FuF-Det’s architecture.

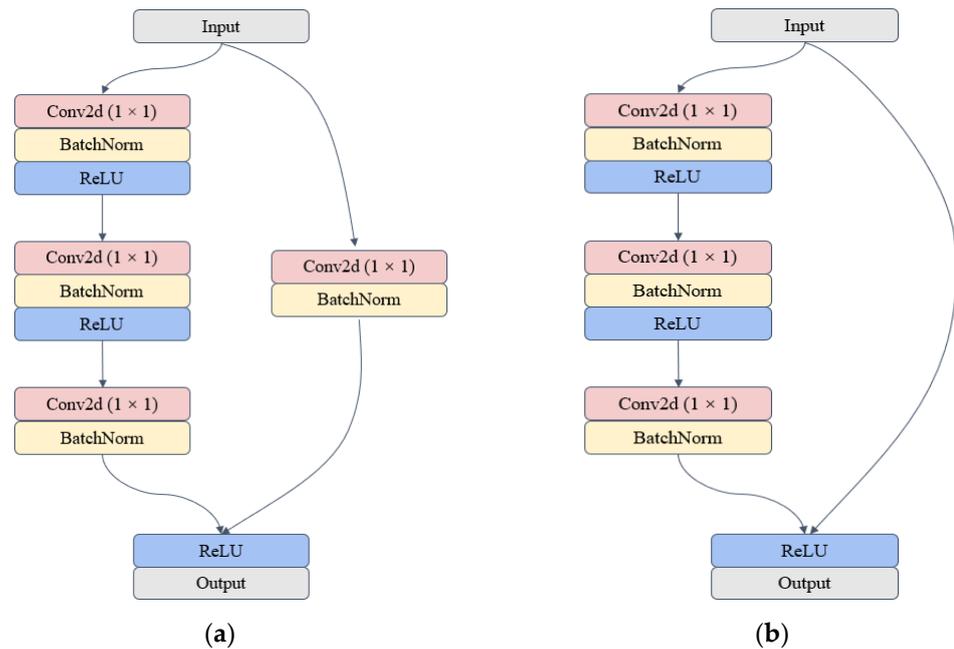


Figure 3. ResNet50 basic block structure: (a) Conv Block; (b) Identity Block.

2.3. Attention-Based Adaptive Fusion Residual Module (AAFRM)

Due to the small size of early forest fire targets, as the network deepens, the encoder may lose the position information of these small fire points, resulting in a low detection rate for the algorithm. The shallow-level features contain rich information about fire point locations, and their combination can provide more reliable detection cues for the algorithm. However, in complex forest environments with small target fire points, the features directly extracted by the network contain a significant amount of forest background information, which is not helpful for fire detection. Therefore, an attention mechanism has been added to the feature fusion module to improve information on the fire point position at both channel and spatial levels.

At the channel level, AAFRM employs efficient channel attention (ECA) [59] to enhance the fire point features. ECA is an efficient attention mechanism proposed in 2020 that adaptively perceives the importance of different channels to improve the algorithm's feature perception capability. In early forest fire images, fire points occupy a small area, and a large amount of redundant background information interferes with the algorithm's extraction of fire point features. Traditional channel attention mechanisms use fully connected layers to change the number of channels and capture correlations among all channels. However, some channels contribute less to the fire point features. Therefore, the features obtained by this method are often affected by the complex forest environment and have weaker representation abilities. In contrast, ECA does not change the number of channels. It calculates correlations between adjacent channels using adaptive one-dimensional convolution, avoiding interference from irrelevant features and improving the information propagation efficiency of the algorithm. The parameter k in ECA represents the range of interaction between channels, which increases with the number of channels. Formula (1) provides the calculation method for k , where $|\cdot|_{odd}$ represents the nearest odd number and γ and b are set to 2 and 1, respectively.

$$k = \psi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{odd} \quad (1)$$

Because focusing only on channel features in the feature fusion structure could cause the algorithm to lose track of where fire points are, AAFRM introduces SimAM [60], which pays attention to both spatial and channel features at the same time. SimAM is a parameter-

free attention mechanism based on similarity. It calculates the similarity and correlation between different features to allocate attention weights to different features, suppress ineffective features, and make the fused features more discriminative and expressive. SimAM evaluates the importance of each neuron by measuring the linear separability between the target neuron and its neighboring neurons. The evaluation criterion is the energy function of the neuron:

$$e_t(\omega_t, b_t, y, x_i) = (y_t - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2 \quad (2)$$

where $\hat{t} = \omega_t + b_t$ and $\hat{x}_i = \omega_i x_i + b_i$ represent the linear transformation of t and x_i , respectively, t denotes the target neuron, and x represents the neighboring neurons. M represents the number of neurons in the channel. A lower energy indicates a more significant difference between the target neuron t and the adjacent neurons, indicating higher importance.

AAFRM incorporates the ECA and SimAM attention mechanisms on top of the direct fusion of shallow and deep features. It also utilizes residual connections to enable the algorithm to focus more on the positional features of early fire points based on encoder features. The AAFRM structure is illustrated in Figure 4.

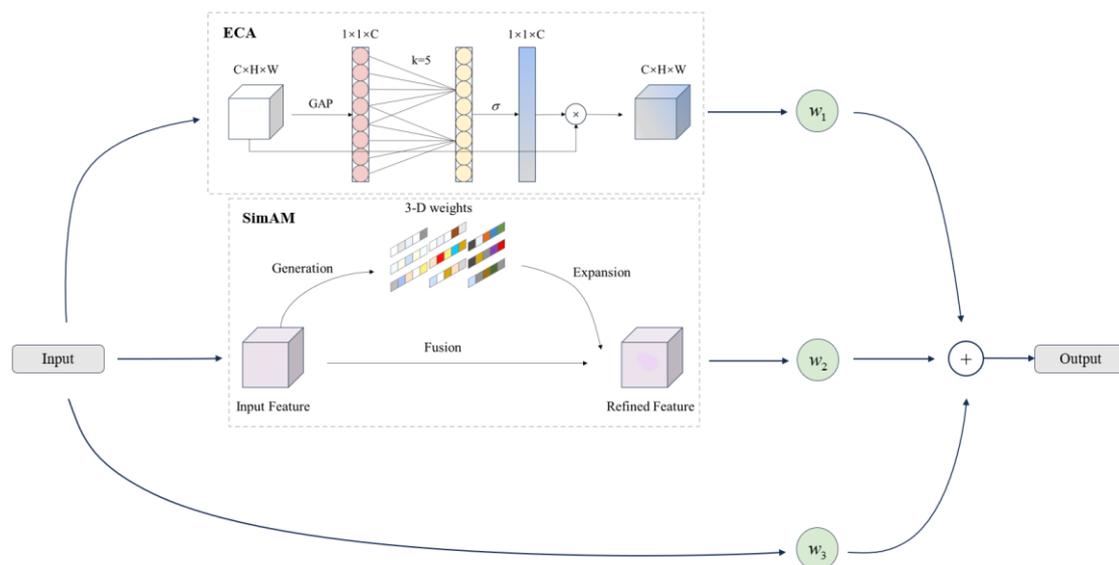


Figure 4. The structure of AAFRM.

First, the input features of the encoder are enhanced by ECA and SimAM, resulting in enriched image features denoted as $eca(x)$ and $simam(x)$, respectively. Then, the network assigns adaptive weights, represented as w_1, w_2, w_3 , to $eca(x)$, $simam(x)$, and x through training. Finally, shallow features are adaptively fused by element-wise addition, enhancing the descriptive capability of features for early fire points. The mathematical expression is:

$$x_{AAFRM} = w_1 \cdot eca(x) + w_2 \cdot simam(x) + w_3 \cdot x \quad (3)$$

2.4. Residual Efficient Channel Attention Block (RECAB)

In the encoder–decoder architecture, the encoder extracts image features that the decoder uses as input for numerous upsampling operations to restore the original data. Common upsampling methods include interpolation, deconvolution, and transpose convolution. Information loss due to padding and convolution operations will always occur, regardless of the method. In the early stages of forest fires in foggy scenes, challenges include small target detection and fog occlusion. Upsampling the features directly from

the encoder output would further lose fire point features. Additionally, the deep features obtained from the encoder contain a significant amount of semantic information, with semantic information being more effectively represented at the channel level. Therefore, in order to preserve detailed information on fire points in foggy conditions, the ECA attention mechanism and residual connection structure are introduced into the transpose convolution used in upsampling, resulting in RECAB. RECAB is designed to efficiently capture and leverage channel-wise dependencies while preserving details of fire points in foggy conditions. By incorporating residual connections, RECAB enables the decoder to recover and enhance the intricate details of fire points during the upsampling process. This module ensures that the decoder effectively restores the fine-grained information of fire points obscured by fog, improving detection performance. The specific structure is illustrated in Figure 5a.

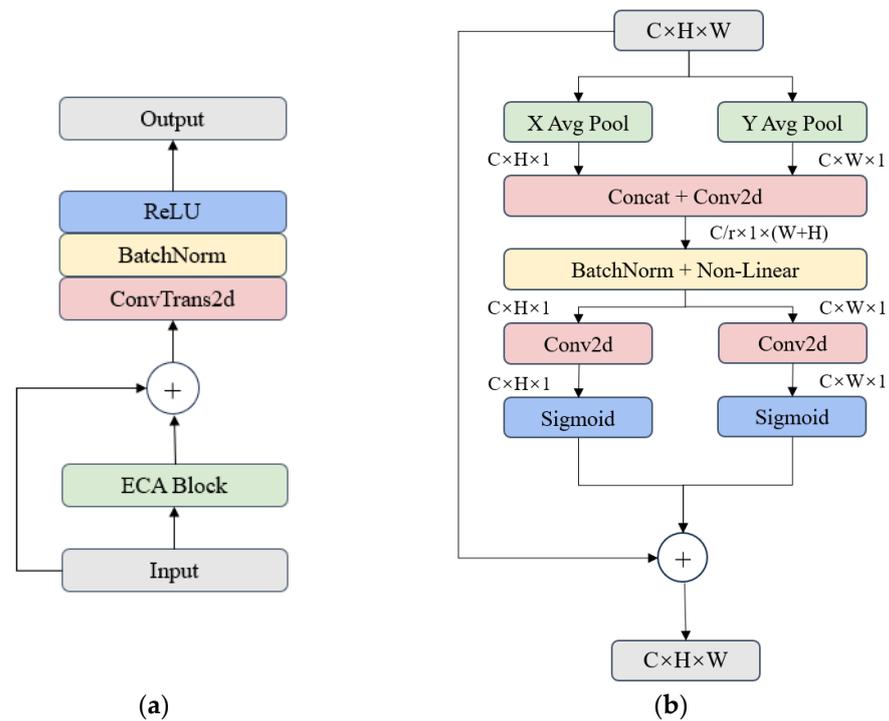


Figure 5. Module structure: (a) RECAB; (b) CA.

In RECAB, the feature x from the previous layer first undergoes the ECA module to obtain the channel-enhanced feature $eca(x)$. The enhanced feature is then fused with the original feature x through a residual connection structure to avoid the loss of detailed features of the fire point targets. Subsequently, a single-layer decoding result is obtained by transpose convolution, BN layer, and ReLU activation function. The mathematical expression is:

$$x_{RECAB} = \sigma(BN(ConvTrans(eca(x) + x))) \quad (4)$$

where $ConvTrans(\cdot)$, $BN(\cdot)$, and $\sigma(\cdot)$ represent the transpose convolution layer, BN layer, and ReLU activation function, respectively.

2.5. CA-Head and Loss Function

2.5.1. CA-Head

When detecting early forest fires, accurately locating fire points can buy firefighters more time. Therefore, the position information of the small target fire points in the image is vital. CA [61] is a variant of the attention mechanism that, based on the channel attention mechanism, applies pooling operations separately in the horizontal and vertical directions of the feature map. This allows the algorithm to capture inter-channel correlations while

preserving accurate positional information, enhancing the representation of target regions. The overall structure of the CA is illustrated in Figure 5b.

FuF-Det is a box-free target detection algorithm that utilizes a center-based approach to locate forest fire points. Center-based target detection consists of three components: heat map, target size, and center point prediction. These components represent, respectively, whether a predicted point contains a fire point, the size of the detection box, and the offset of the predicted center point on the feature map. In order to emphasize the position information of the targets during the detection process, CA is incorporated into the box-free detection head to obtain the CA-Head detection module. The addition of CA allows the algorithm to associate image features with target position information, enabling it to distinguish between foggy backgrounds and fire points in forest environments, thus improving the accuracy of detection boxes. The overall structure of the CA-Head is illustrated in Figure 6.

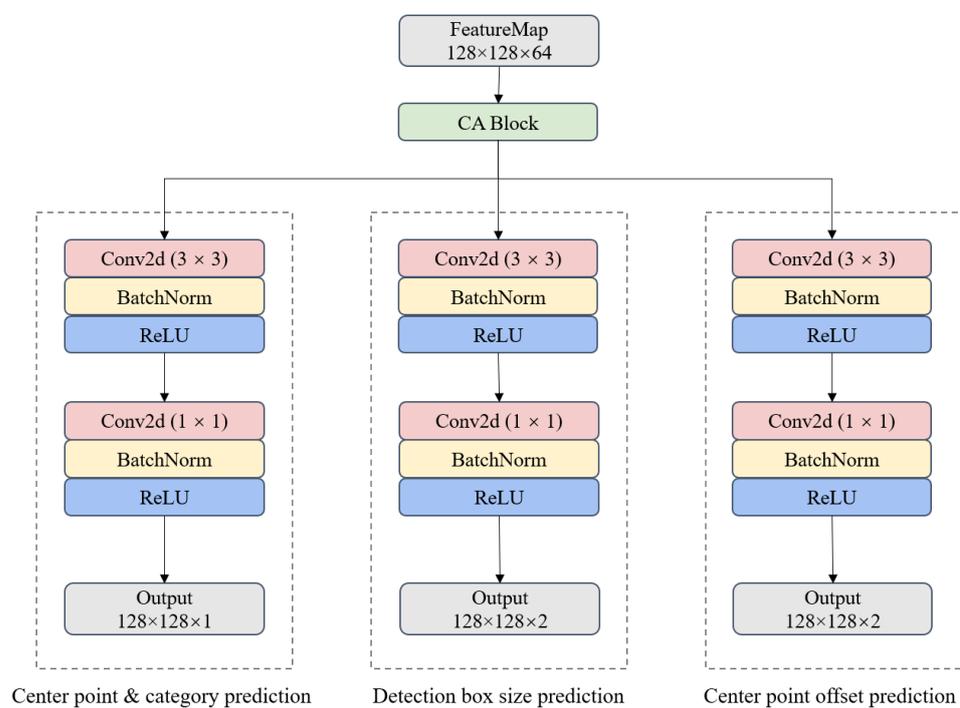


Figure 6. The structure of CA-Head.

2.5.2. Loss Function

Compared to common anchor-based algorithms, anchor-free detection based on center points does not require pre-defined anchor box sizes. Instead, it utilizes target center points as anchor box representatives. It uses image features of the center points to regress, determine whether a target is a fire point, and locate fire positions. CA-Head predicts the heat map, target center points, and target sizes through three independent convolutional modules. Therefore, the loss function consists of three components: category loss, center point offset loss, and detection box size loss. First, the predicted category loss L_k for the k th image is calculated using the concept of focal loss [62].

$$L_k = \frac{1}{N} \sum_{xyc} \begin{cases} (1 - \hat{Y}_{xyc})^\alpha \log(\hat{Y}_{xyc}) & \text{if } Y_{xyc} = 1 \\ (1 - \hat{Y}_{xyc})^\beta (\hat{Y}_{xyc})^\alpha \log(1 - \hat{Y}_{xyc}) & \text{others} \end{cases} \quad (5)$$

where \hat{Y}_{xyc} represents the predicted results of the algorithm and Y_{xyc} represents the ground truth labels. N denotes the number of key points predicted by the algorithm in the image, and the subscript xyc represents the detected sample. The hyperparameters α and β are set to 2 and 4, respectively.

The center point offset loss function L_{off} and the detection box size loss function L_{size} are both calculated using the L1 loss function:

$$L_{off} = \frac{1}{N} \sum_p \left| \hat{o}_{\tilde{P}} - \left(\frac{p}{R} - \tilde{P} \right) \right| \quad (6)$$

$$L_{size} = \frac{1}{N} \sum_{k=1}^N \left| \hat{S}_{pk} - S_k \right| \quad (7)$$

In the center point offset loss function, p represents the actual coordinates of the fire points in the image. R represents the downsampling factor in the detection head, with $R = 4$ in this paper. $\hat{O}_{\tilde{P}}$ represents the predicted offset of the fire points. \tilde{P} represents the predicted center point coordinates of the fire points. $\frac{p}{R} - \tilde{P}$ represents the actual offset of the center point of the fire points. In the detection box size loss function, S_k represents the actual size of the detection box for fire points. \hat{S}_{pk} represents the size of the detection box obtained by regression calculation. Therefore, the overall loss function of FuF-Det is the sum of L_{off} , L_{size} , and L_k multiplied by their respective coefficients:

$$L_{sum} = L_k + \lambda_{size} L_{size} + \lambda_{off} L_{off} \quad (8)$$

where $\lambda_{size} = 0.1$, $\lambda_{off} = 1$.

3. Experiment and Results

3.1. Dataset

The dataset used in the experiments for early forest fire detection in foggy conditions was constructed based on the publicly available dataset FLAME2 [57]. The creators of the dataset mention that the FLAME2 image is of smoke. However, FLAME2 is chosen as a dataset for fire detection under fog. This is because when a fire occurs, there is a mixture of smoke and fog around the fire point. We believe that training the model with FLAME2 can improve the generalization of the model and its applicability in real-world scenarios. FLAME2 contains 52,287 visible mild forest fire images captured by UAVs. Since these images are from successive UAV frames, they exhibit high similarity between successive images. Therefore, data cleaning was performed on the images first. Specifically, histogram features were used to measure image similarity. Subsequent images were removed if the similarity between an image and its four subsequent images exceeded a predefined threshold (set to 0.98 in this paper). After data cleaning, the dataset consists of 2349 254×254 pixel images, including 984 non-fire images and 1365 fire images with fog. Data augmentation was performed on the cleaned dataset to avoid overfitting due to the limited amount of training data, including adding noise, cropping, translation, rotation, and flipping. The augmented dataset contains 14,094 images. The augmented dataset maintains a ratio of 984/1365 for non-fire/fire images. Subsequently, the LabelImg target detection annotation software was used to annotate the fire in the dataset. Finally, a random selection of 11,415 images was selected as the training set, 1269 as the validation set, and 1410 as the test set. Figure 7 shows some sample images.

3.2. Model Evaluation

Precision, Recall, Average Precision (AP), F1 score, and Frames Per Second (FPS) were adopted as the evaluation metrics for our algorithm. Precision indicates the probability that the detected targets belong to fires. Recall represents the probability of the algorithm detecting all actual fire targets. The F1 score combines Precision and Recall, serving as their harmonic mean. AP represents the area under the P-R curve for fires, and in this paper, the detection performance of the algorithm was comprehensively evaluated using

AP@0.5. The formulas for calculating Precision, Recall, F1 score, and AP are shown in Equations (9)–(12):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

$$F1 = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

$$AP = \int P(r)dr \quad (12)$$

TP (True Positive) represents the fire targets and is predicted as fire. *FP* (False Positive) represents non-fire targets but is predicted to be fire. *FN* (False Negative) represents targets that are actually fire but predicted as non-fire.

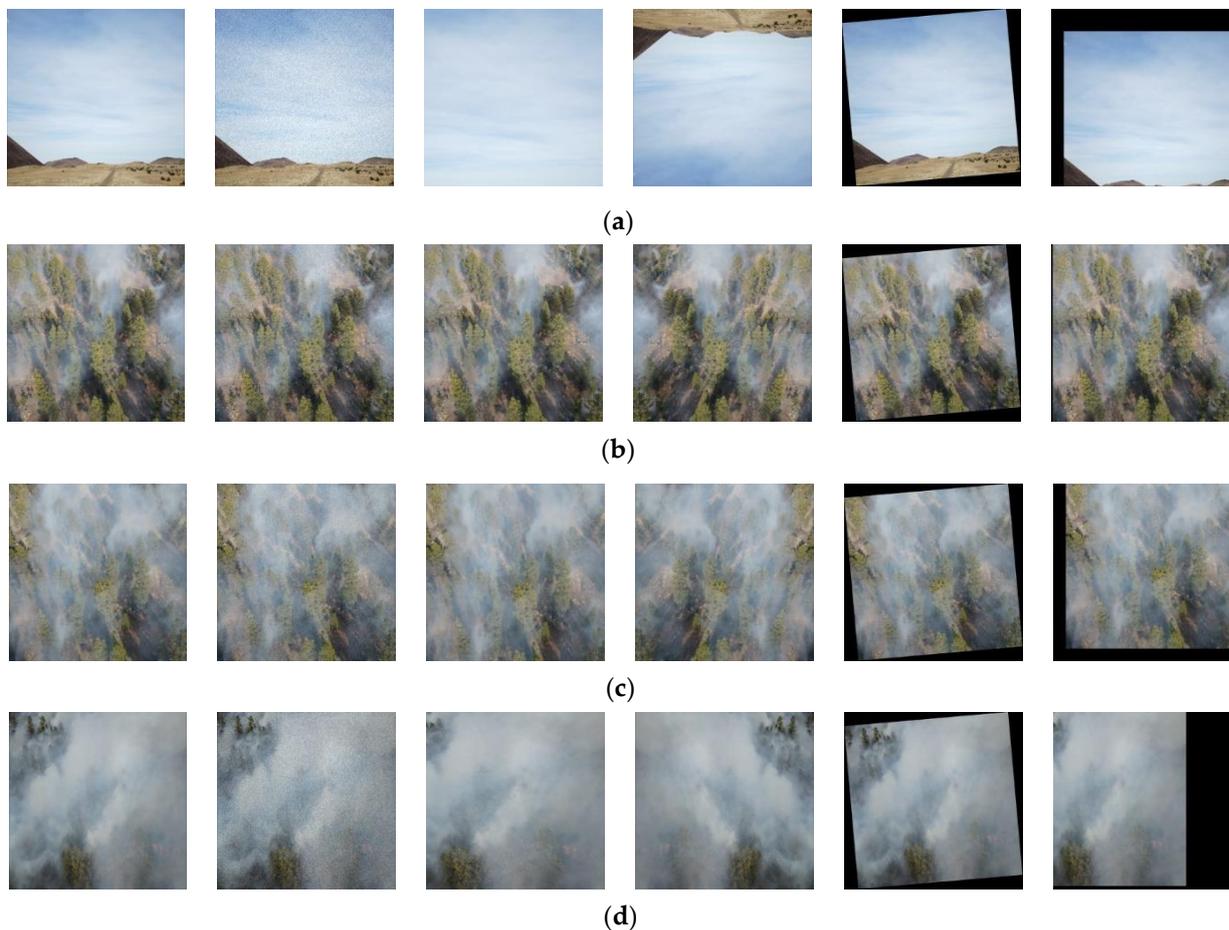


Figure 7. Example dataset sample image: (a–d) No fog, mild, moderate, and severe fog scenes, respectively. (Columns 1 to 6 are the original image, the noise-added, cropped, flipped, rotated, and shifted image).

3.3. Training

3.3.1. Experimental Environment

The hardware environment for the experiment is a Windows 11 system with a 12th Gen Intel (R) Core (TM) i7-12700 CPU @ 2.10 GHz, 32 GB RAM, and an NVIDIA GeForce RTX 3060 GPU with 12 GB of VRAM. The software environment includes Python 3.9, PyTorch 1.13.1, and the PyCharm editor 2022.3.2.

3.3.2. Training Parameter Settings

The encoder's role is to extract deep features from images captured by UAVs. For different detection tasks, the features of the targets are similar. Therefore, firstly, the pre-trained weights of ResNet50 trained on the VOC dataset were used during the training process to perform frozen training for 50 epochs, which allowed the training to focus on feature fusion, decoder, and detection head. After 50 epochs, the encoder weights for the overall detection task were finetuned during unfrozen training. The training process uses the Adam optimizer [63] with an initial learning rate of 5×10^{-4} and a minimum learning rate of 5×10^{-6} . To avoid the problem of becoming stuck in local optima during training, cosine annealing decay is used to adjust the learning rate. Specific parameters for frozen and unfrozen training are shown in Table 2.

Table 2. Parameter settings for frozen and unfrozen training.

	Training Epoch	Batch Size	Input Size	Optimizer	Momentum
Freeze training	1–50	16	512 × 512	Adam	0.9
Unfreeze training	51–200	8			

3.4. Experimental Results

3.4.1. Comparison with Other Target Detection Algorithms

To validate the advantages of FuF-Det in detecting early forest fires in foggy scenes, we compared it with several commonly used target detection algorithms. Eight detection algorithms were included in the comparison: the two-stage detection algorithm Faster R-CNN [64], the one-stage detection algorithms YOLOv3 [65], YOLOv4 [66], YOLOv5_m, YOLOv7 [67], the anchor-free target detection algorithms YOLOX [68], YOLOv8_s [69], and CenterNet [70]. All algorithms were trained on the augmented dataset in the same experimental environment. Table 3 presents the detection results of these eight target detection algorithms, including FuF-Det, on the test set.

Table 3. Experimental results of comparative experiments.

	Method	AP@0.5	Pre	Rec	F1	FPS
Anchor-based	Faster R-CNN (ResNet50)	12.57%	26.22%	25.28%	0.26	11
	YOLOv3	66.26%	74.87%	69.63%	0.72	36
	YOLOv4	64.05%	81.87%	50.84%	0.63	32
	YOLOv5_m	80.25%	87.66%	69.02%	0.77	33
	YOLOv7	65.56%	85.30%	44.21%	0.58	26
Anchor-free	YOLOX	81.23%	87.61%	73.69%	0.80	21
	YOLOv8_s	86.46%	92.46%	73.36%	0.82	58
	CenterNet (ResNet50)	76.72%	87.58%	68.55%	0.77	30
	FuF-Det (Ours)	86.52%	91.87%	78.69%	0.85	20

The analysis in Table 3 shows that FuF-Det outperforms the other eight target detection algorithms in detecting early forest fires in foggy scenes. The two-stage Faster R-CNN detection algorithm performs the worst in this task, with an average accuracy of only 12.57% on the test set, because the two-stage target detection algorithms generate a series of candidate boxes and then classify and locate these candidates. However, early forest fire targets are small, and the presence of fog affects the selection of candidate regions, leading to inaccurate candidate box generation and missing and false detections. Additionally, generating candidate boxes reduces the algorithm's detection speed, which is why Faster R-CNN has the lowest real-time performance among the nine algorithms. In contrast, FuF-

Det belongs to the one-stage detection algorithms, achieving the highest *AP*, *Recall*, and *F1* score on the test set, reaching 86.52%, 78.69%, and 0.85, respectively. Compared to other one-stage target detection algorithms, including YOLOv3, YOLOv4, YOLOv5_m, YOLOv7, YOLOX, YOLOv8_s, and CenterNet, FuF-Det improves the *AP* by 20.26%, 22.47%, 6.27%, 20.96%, 5.29%, 0.06%, and 9.8%, respectively. In terms of *Precision*, YOLOv8_s achieves 92.46%, which is only 0.59% higher than FuF-Det. Moreover, with a slight difference in *Precision*, FuF-Det achieves a *Recall* of 78.69%, which is an improvement of 5.33% compared to YOLOv8_s. In early forest fire detection tasks, it is crucial for an algorithm to maintain high *Precision* while detecting fires with a high detection rate. Therefore, the slightly lower *Precision* compared to YOLOv8_s does not affect FuF-Det's superior performance in detecting early forest fires in foggy scenes among the nine detection algorithms. In terms of real-time performance, FuF-Det sacrifices some detection speed due to the introduction of the AAFRM feature fusion module, which improves the issue of early fire point localization information loss. The FPS of FuF-Det is 20, which is the slowest of all the one-stage detection algorithms. However, it is still sufficient to meet the real-time requirements of UAV detection.

Eight one-stage target detection algorithms were used to find early forest fires when there was no fire and different levels of fog (mild, moderate, and severe fog). The results are shown in Figures 8–11, along with heat maps showing the results. From the results, it can be observed that, of the three scenarios listed, except for the proposed FuF-Det, the other target detection algorithms exhibit varying degrees of false detection and missed detection. The presence of red-colored targets, which in the images correspond to people wearing red clothing, is the main reason for the incorrect detection problem. It is worth noting that, in the severe fog scenario, apart from FuF-Det, none of the other target detection algorithms were able to detect fire points.

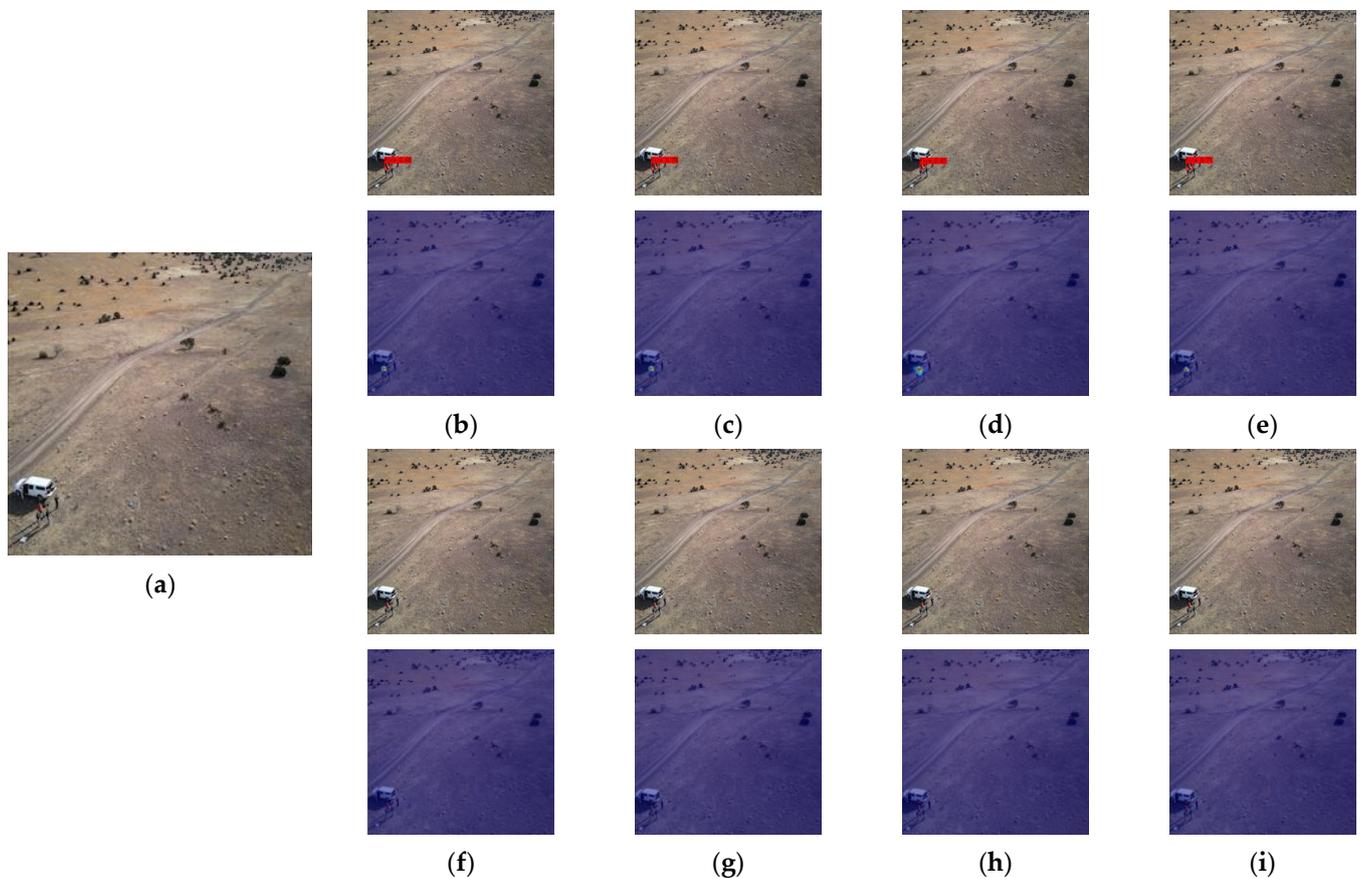


Figure 8. Detection result in the non-fire scenario: (a–i) indicates ground truth, the detection results of YOLOv3, YOLOv4, YOLOv5_m, YOLOv7, YOLOX, YOLOv8_s, CenterNet and FuF-Det.

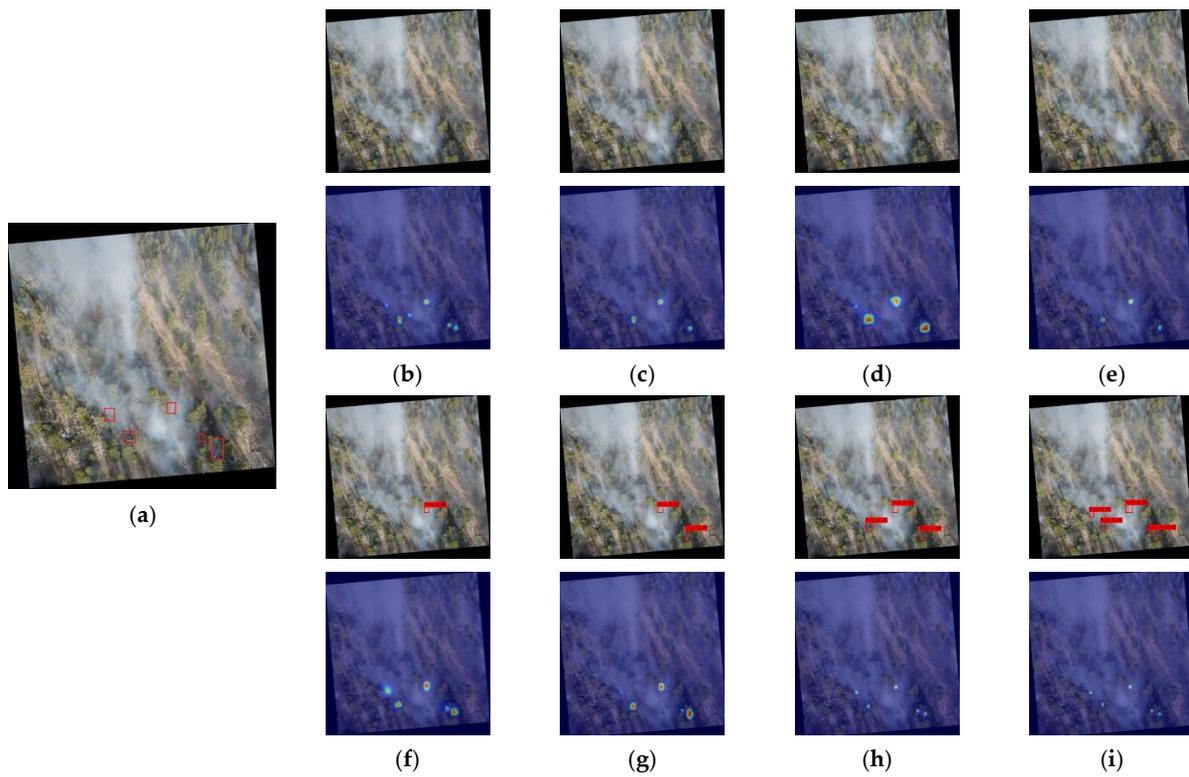


Figure 9. Detection results in mild fog scenarios: (a–i) indicates ground truth (red boxes), the detection results of YOLOv3, YOLOv4, YOLOv5_m, YOLOv7, YOLOX, YOLOv8_s, CenterNet and FuF-Det.

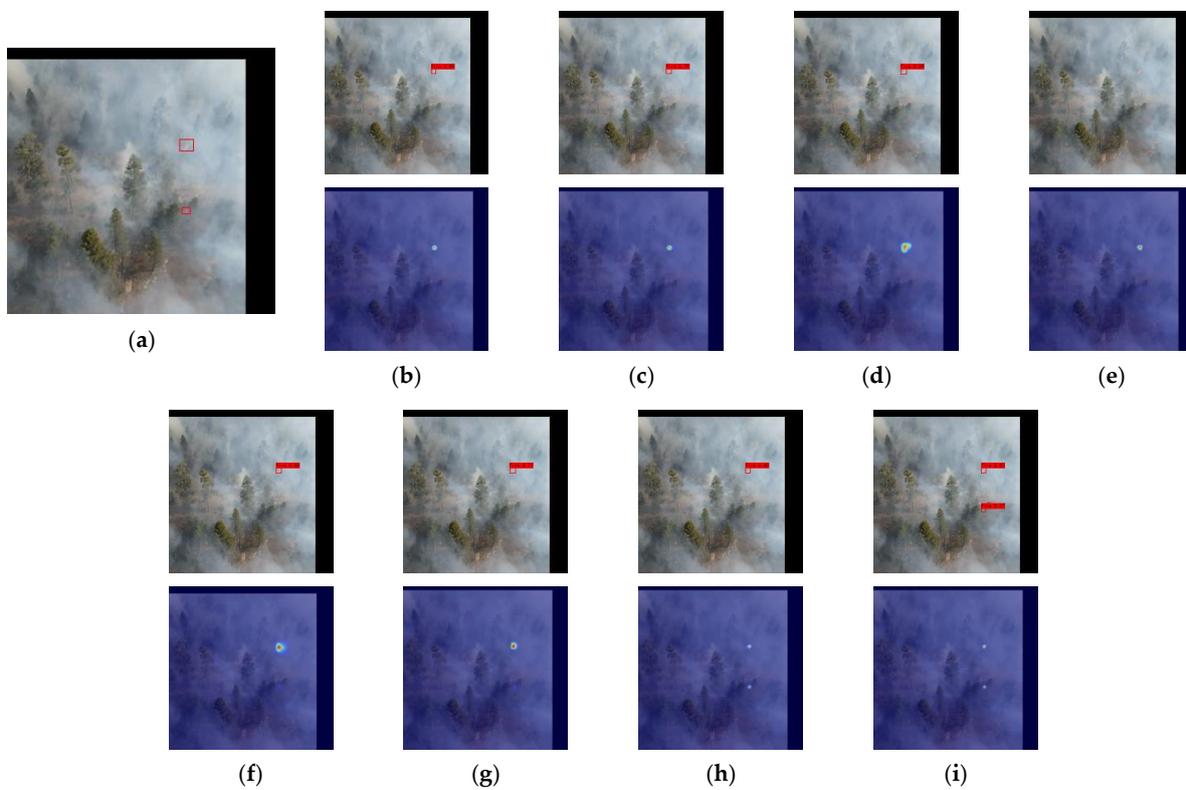


Figure 10. Detection results in moderate fog scenarios: (a–i) indicates ground truth (red boxes), the detection results of YOLOv3, YOLOv4, YOLOv5_m, YOLOv7, YOLOX, YOLOv8_s, CenterNet and FuF-Det.

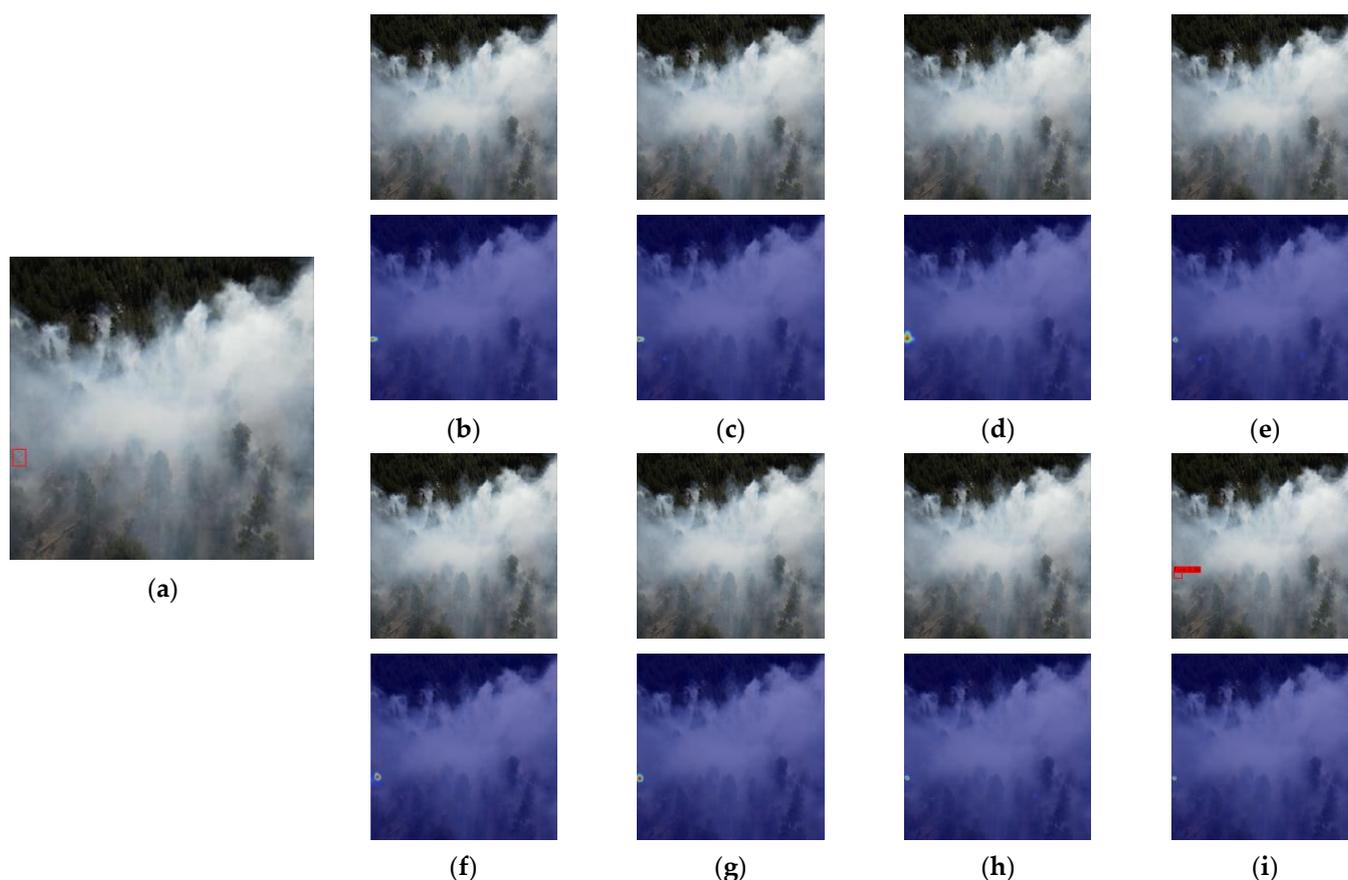


Figure 11. Detection results in severe fog scenarios: (a–i) indicates ground truth (red box), the detection results of YOLOv3, YOLOv4, YOLOv5_m, YOLOv7, YOLOX, YOLOv8_s, CenterNet and FuF-Det.

3.4.2. Ablation Experiments

Ablation experiments were conducted to compare and analyze the different modules to demonstrate the effectiveness of the AAFRM feature fusion unit, RECAP decoder unit, and CA-Head detection module proposed in this paper. The results of ablation experiments are shown in Table 4. Experiment 1 serves as the baseline, while CenterNet with ResNet50 is used as the backbone network. Experiment 2, based on Experiment 1, involves training with an augmented dataset. As shown in the results, Experiment 2 outperforms Experiment 1 in all metrics. The augmented dataset provides a more prosperous and diverse training sample, enabling the algorithm to learn the invariance and features of the targets, thus improving the robustness and generalization ability of the algorithm. Comparing the results of Experiment 3 and Experiment 2, including the AAFRM feature fusion unit increases the detection rate of small fire points by 9.25% because AAFRM adaptively enhances the position information of fire points in the feature maps at both spatial and channel levels. The fused feature maps contain semantic information describing the details of fire points and include position information of small fire points in the global context, making it easier for the algorithm to locate early fire positions. RECAP, as the encoder unit, utilizes residual connections and attention mechanisms to address the issue of fire point information loss caused by upsampling in the encoder, which ensures that the encoder's output contains richer semantic information about fire points, facilitating the algorithm in distinguishing actual fire points from fire-like targets. Therefore, Experiment 4, compared to Experiment 3, shows a more significant improvement in Precision (3.04%) but a smaller improvement in Recall (5.52%). In Experiment 5, the CA-Head is added before the detection head. By focusing on the position information of fire points in the image, CA-Head enables the algorithm to accurately predict the positions and bounding

boxes of fire points. Additionally, CA helps the algorithm associate the features of fire points with their respective positions in the image, allowing better capture of contextual information and spatial relationships of fire points, thus enhancing the recognition and classification capabilities of CA-Head for fire points. Compared to Experiment 2, CA-Head improves the algorithm's AP, Precision, and Recall by 6.01%, 2.82%, and 5.24%, respectively. Experiments 6 to 8 combine AAFRM, RECAB, and CA-Head in pairs, and the results show that adding each corresponding module improves the detection capability of the algorithm. It is worth noting that the introduction of RECAB and CA modules does not affect the detection speed of the algorithm. In other words, the proposed RECAB unit and CA-Head can ensure real-time detection while significantly improving the detection capability of the algorithm. Experiment 8, which incorporates both RECAB and CA-Head modules, even improves detection speed, indicating that the simultaneous addition of both modules provides the algorithm with more robust stability and generalization ability, which is of great value for future research. Moreover, Experiment 6, which includes both AAFRM and RECAB modules, achieves the highest recall, indicating that the proposed idea of self-enhancing the encoder features before fusing them with decoder features, as well as the feature enhancement operation in the decoder, effectively cooperates in the overall algorithm, allowing the algorithm to detect more small fire points, which also provides direction for future research.

Table 4. Results of ablation experiments.

Experiment	Data Augmentation	AAFRM	RECAB	CA-Head	AP@0.5	Pre	Rec	FPS
1					70.08%	87.43%	46.78%	26
2	✓				76.72%	87.58%	68.55%	30
3	✓	✓			83.48%	89.81%	77.80%	21
4	✓		✓		82.50%	90.62%	74.07%	30
5	✓			✓	82.73%	90.40%	73.79%	30
6	✓	✓	✓		85.36%	91.32%	79.63%	21
7	✓	✓		✓	84.67%	91.29%	79.30%	21
8	✓		✓	✓	83.25%	90.94%	74.67%	33
FuF-Det	✓	✓	✓	✓	86.52%	91.87%	78.69%	20

3.4.3. Missed Detection Analysis

Although FuF-Det shows superior fire detection capability under foggy conditions compared to mainstream target detection algorithms, it still inevitably encounters the issue of missed detection in practical scenarios. Figure 12 illustrates three instances of missed detections. In the first column, the images represent the algorithm's detection results, with red bounding boxes indicating the targets detected by the algorithm and green bounding boxes representing the missed fire point targets. Analysis reveals that FuF-Det's detection capability is compromised when the extremely small fire points and the fog are particularly dense. Furthermore, the coexistence of fog and forest background contributes to the algorithm's missed detection problem, which could be attributed to three factors. First, the low resolution of the dataset reduces the saliency of small fire points in the images. Second, after data augmentation operations, the clarity of the images is reduced compared to the original images, further decreasing the saliency of small fire points and increasing the difficulty of their detection by the algorithm. Additionally, the limited amount of feature information of small fire points is further weakened when fog and forest background coexist, compromising the algorithm's ability to extract these features. In conclusion, to enhance the practical application effectiveness of FuF-Det, further improvements are required in terms of image quality during data collection and the algorithm's feature extraction capability for small targets in complex scenes.



Figure 12. FuF-Det missed detection results (the green box and the red box indicate the missed and detected fire points, respectively).

4. The Detection Effect of FuF-Det in Different Scenes

The dataset used in the experiment in this paper is all from FLAME2, which contains a relatively simple scene, and it is difficult to see the adaptability of FuF-Det in different fire scenes from the detection results. In order to further explore the detection effect of FuF-Det in different scenes, fire detection experiments for snowy forest scenes and non-forest fire scenes are designed.

4.1. Snowy Forest Scene

The FLAME dataset [71] provides aerial UVA-based images of snowy forest fires. Compared to the dataset used in this paper, the images in it are not obscured by fog, but there is a problem of vegetation obscuring the fire point. FuF-Det detection results in the snow scene are shown in Figure 13. In the original image, the red box represents where the actual fire point exists.

As shown in Figure 13, FuF-Det is also suitable for the task of fire detection in snowy forests and can effectively deal with the problem that the fire point is obscured by vegetation. To a certain extent, this indicates that FuF-Det is adaptable to changes in weather and fire point occlusions in forest scenes.



Figure 13. *Cont.*



Figure 13. Detection results of FuF-Det on FLAME dataset (the left column is the original image with red boxes indicating the fire points, and the right column is the FuF-Det detection result).

4.2. Non-Forest Fire Scenes

In order to further explore the possibility of FuF-Det application in other fire scenes, fire images in non-forest scenes were collected from the network and detected by FuF-Det. The test results are shown in Figure 14.

According to Figure 14a,b, when the detection environment is no longer a forest, FuF-Det can still detect the fire point with high accuracy. Even in the case of more and more dispersed fire points presented in (b). However, Figure 14c,d shows where FuF-Det is still lacking. When there are both large and small fire points in the image, FuF-Det usually gives priority to the location of small fire points and ignores the detection of large fire points. This problem arises in the forest scene as well as in other scenes. This is because, in the FuF-Det training process, the training set contains small fire points, resulting in more fire point features learned by FuF-Det from small fire points. In future work, the images in the dataset will be optimized, increasing the number of large-size fire point images, and improving the adaptability of FuF-Det in large fire scenes. Table 5 summarizes the adaptability of FuF-Det when different conditions change in the fire scene.

Table 5. A summary of the adaptability of FuF-Det.

Variable	Season	Obstructions	Scene	Fire Point Size
The adaptability of FuF-Det	✓	✓	✓	×



(a) Building fire

Figure 14. Cont.



(b) Mountain fire



(c) Example of large size fire point missing detection



(d) Example of large-size fire point missing detection

Figure 14. Detection results of FuF-Det in non-forest scenes (the left column is the original image, and the right column is the FuF-Det detection result).

5. Conclusions

In order to detect early forest fire points obscured by fog during UAV inspections, this paper proposes the FuF-Det algorithm for early forest fire detection under foggy conditions. FuF-Det is an anchor-free detection algorithm based on an encoder–decoder structure. First, ResNet50 is used as the encoder, taking advantage of residual connections to extract deeper features of early forest fire points. Second, RECAP is designed as the decoder unit, which combines residual structures and ECA to effectively address the issue of feature loss for fire points caused by upsampling in the encoder. Furthermore, to improve the detection rate of small fire points, the AAFRM is designed as a feature fusion unit to enhance the position information of fire points at both channel and spatial levels. Finally, CA is introduced before the detection head to obtain the CA-Head module, which helps the algorithm predict whether the target is a fire point while considering its positional information. This enables a more accurate generation of detection boxes for small fire points.

The ablation experiments show that the proposed AAFRM, RECAB, and CA-Head modules can effectively detect early forest fires in foggy scenes. Moreover, the experimental results show that the one-stage target detection algorithms have a better detection effect on the early forest fire point under fog. YOLOv5_m, YOLOX, YOLOv8_s, and the proposed FuF-Det all showed good fire point detection performance. However, Figure 9d,f,g show that YOLOv5_m, YOLOX, YOLOv8_s have a large range of target perception and cannot accurately locate early fire points. In addition, Figure 10d,f,g show that these models have limited ability to extract features from fire spots with severe occlusion, which can easily lead to the problem of missing fire points. In contrast, FuF-Det demonstrated better detection capabilities with an $AP@0.5$ of 86.52% and a fire spot detection rate of 78.69%, which can effectively deal with the detection tasks of small fire points and fog occlusion in the early stage. Finally, fire images in different scenes are tested by FuF-Det. The results show that FuF-Det can maintain good adaptability when the forest season, the obstruction, and the fire scene change.

In the future, further studies on FuF-Det will be conducted. First, for the dataset used in the study, more non-fire, different seasons, different forest types, etc. images will be added to improve the generalization of the model so that it can better adapt to the scene of non-fire, forest season, and vegetation type changes encountered in the detection process. We will also find specific indicators to quantify the degree of fog occlusion and refine the images in the dataset to continue to study the influence of different occlusion degrees on the detection ability of FuF-Det. For the method, aiming at the problem of the AAFRM feature fusion unit reducing the detection speed and the problem of missing detection caused by a complex forest background, the model or module, according to the results, will also be optimized to help the UAVs accurately and quickly detect early forest fires under fog conditions. Finally, the possibility of applying FuF-Det to the detection of other forest resources, such as the monitoring of wildlife biomass, forest pests, and diseases, will be explored, to promote the application of FuF-Det in forest resource monitoring.

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

Funding: This research was funded by the National Nature Science Founding of China grant number 61573183.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the need for future work.

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

Abbreviations

VAM	Value conversion-Attention mechanism Module
BiFPN	Bidirectional Feature Pyramid Network
YCbCr	Luminance, Colour-difference of blue, Colour-difference of red
HSV	Hue, Saturation, Value
HSL	Hue, Saturation, Lightness
HWB	Hue, Whiteness, Blackness
MS-FRCNN	Multi-Scale Faster RCNN Model
R-CNN	Region-CNN
MTL-FFDET	Multi-Task Learning-Based Model for Forest Fire Detection
GXLD	GhostNet-YOLOX-L-Light-Defog

References

1. Pan, Y.; Birdsey, R.A.; Fang, J.; Houghton, R.; Kauppi, P.E.; Kurz, W.A.; Phillips, O.L.; Shvidenko, A.; Lewis, S.L.; Canadell, J.G.; et al. A Large and Persistent Carbon Sink in the World's Forests. *Science* **2011**, *333*, 6045. [CrossRef] [PubMed]
2. Haynes, K.; Short, K.; Xanthopoulos, G.; Viegas, D.; Ribeiro, L.M.; Blanchi, R. Wildfires and WUI fire fatalities. In *Encyclopedia of Wildfires and Wildland/Urban Interface (WUI) Fires*; Manzello, S.L., Ed.; Springer: Berlin/Heidelberg, Germany, 2020; pp. 1–16.
3. Whittaker, J.; Handmer, J.; Mercer, D. Vulnerability to bushfires in rural Australia: A case study from East Gippsland, Victoria. *J. Rural Stud.* **2012**, *28*, 161–173. [CrossRef]
4. Forest Monitoring, Land Use & Deforestation Trends. Global Forest Watch. Available online: <https://www.globalforestwatch.org/> (accessed on 14 September 2023).
5. Baijnath-Rodino, J.A.; Li, S.; Martinez, A.; Kumar, M.; Quinn-Davidson, L.N.; York, R.A.; Banerjee, T. Historical seasonal changes in prescribed burn windows in California. *Sci. Total Environ.* **2022**, *836*, 155723. [CrossRef] [PubMed]
6. Swain, D.L.; Abatzoglou, J.T.; Kolden, C.; Shive, K.; Kalashnikov, D.A.; Singh, D.; Smith, E. Climate change is narrowing and shifting prescribed fire windows in western United States. *Commun. Earth Environ.* **2023**, *4*, 340. [CrossRef]
7. Dampage, U.; Bandaranayake, L.; Wanasinghe, R.; Kottahachchi, K.; Jayasanka, B. Forest fire detection system using wireless sensor networks and machine learning. *Sci. Rep.* **2022**, *12*, 46. [CrossRef]
8. Sinha, D.; Kumari, R.; Tripathi, S. Semisupervised Classification Based Clustering Approach in WSN for Forest Fire Detection. *Wirel. Pers. Commun.* **2019**, *109*, 2561–2605. [CrossRef]
9. Yu, L.; Wang, N.; Meng, X. Real-time forest fire detection with wireless sensor networks. In Proceedings of the 2005 International Conference on Wireless Communications, Networking and Mobile Computing, Wuhan, China, 26 September 2005; pp. 1214–1217.
10. Kang, Y.J.; Jang, E.; Im, J.; Kwon, C.G. A deep learning model using geostationary satellite data for forest fire detection with reduced detection latency. *GIScience Remote Sens.* **2022**, *59*, 2019–2035. [CrossRef]
11. Fernandes, A.M.; Utkin, E.; Lavrov, A.V.; Vilar, R.M. Development of neural network committee machines for automatic forest fire detection using lidar. *Pattern Recognit.* **2004**, *37*, 2039–2047. [CrossRef]
12. Chen, S.K.; Cao, Y.C.; Feng, X.Q.; Lu, X.B. Global2Salient: Self-adaptive feature aggregation for remote sensing smoke detection. *Neurocomputing* **2021**, *466*, 202–220. [CrossRef]
13. Zheng, Y.; Zhang, G.; Tan, S.Q.; Yang, Z.G.; Wen, D.X.; Xiao, H.S. A forest fire smoke detection model combining convolutional neural network and vision transformer. *Front. For. Glob. Change* **2023**, *6*, 1136969. [CrossRef]
14. Li, X.L.; Song, W.G.; Lian, L.P.; Wei, X.G. Forest Fire Smoke Detection Using Back-Propagation Neural Network Based on MODIS Data. *Remote Sens.* **2015**, *7*, 4473–4498. [CrossRef]
15. Sudhakar, S.; Vijayakumar, V.; Sathiy Kumar, C.; Priya, V.; Ravi, L.; Subramaniaswamy, V. Unmanned Aerial Vehicle (UAV) based Forest Fire Detection and monitoring for reducing false alarms in forest-fires. *Comput. Commun.* **2020**, *149*, 1–16. [CrossRef]
16. Zhan, J.L.; Hu, Y.W.; Zhou, G.X.; Wang, Y.F.; Cai, W.W.; Li, L.J. A high-precision forest fire smoke detection approach based on ARGNet. *Comput. Electron. Agric.* **2022**, *196*, 106874. [CrossRef]
17. Zhang, L.; Wang, M.Y.; Ding, Y.H.; Bu, X.F. MS-FRCNN: A Multi-Scale Faster RCNN Model for Small Target Forest Fire Detection. *Forests* **2023**, *14*, 616. [CrossRef]
18. Zheng, H.T.; Dembélé, S.; Wu, Y.X.; Liu, Y.; Chen, H.L.; Zhang, Q.J. A lightweight algorithm capable of accurately identifying forest fires from UAV remote sensing imagery. *Front. For. Glob. Change* **2023**, *6*, 1134942. [CrossRef]
19. Alkhatib, A.A.A. A Review on Forest Fire Detection Techniques. *Int. J. Distrib. Sens. Netw.* **2014**, *10*, 597368. [CrossRef]
20. 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]
21. Cruz, H.; Gualotuña, T.; Pinillos, M.; Marcillo, D.; Jácome, S. Machine Learning and Color Treatment for the Forest Fire and Smoke Detection Systems and Algorithms, a Recent Literature Review. In *Artificial Intelligence, Computer and Software Engineering Advances: Proceedings of the CIT 2020, Quito, Ecuador, 26–30 October 2020*; Springer: Berlin/Heidelberg, Germany; pp. 109–120.
22. Moulianitis, V.C.; Thanellas, G.; Xanthopoulos, N.; Aspragathos, N.A. Evaluation of UAV Based Schemes for Forest Fire Monitoring. In *Advances in Service and Industrial Robotics: Proceedings of the 27th International Conference on Robotics in Alpe-Adria Danube Region (RAAD 2018), Patras, Greece, 6–8 June 2018*; Springer: Berlin/Heidelberg, Germany; pp. 143–150.
23. Ko, B.C.; Kwak, J.Y.; Nam, J.Y. Wildfire smoke detection using temporospatial features and random forest classifiers. *Opt. Eng.* **2012**, *51*, 017208. [CrossRef]
24. Prema, C.E.; Vinsley, S.S.; Suresh, S. Multi Feature Analysis of Smoke in YUV Color Space for Early Forest Fire Detection. *Fire Technol.* **2016**, *52*, 1319–1342. [CrossRef]
25. Peng, Y.S.; Wang, Y. Real-time Forest smoke detection using hand-designed features and deep learning. *Comput. Electron. Agric.* **2019**, *167*, 105029. [CrossRef]
26. Sun, X.F.; Sun, L.P.; Huang, Y.L. Forest fire smoke recognition based on convolutional neural network. *J. For. Res.* **2021**, *32*, 1921–1927. [CrossRef]
27. Almeida, J.S.; Huang, C.X.; 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]
28. Sathishkumar, V.E.; Cho, J.; Subramanian, M.; Naren, O.S. Forest fire and smoke detection using deep learning-based learning without forgetting. *Fire Ecol.* **2023**, *19*, 9. [CrossRef]

29. Wu, J.; Huang, R.L.; Xu, Z.Y.; Han, N. Forest fire smog feature extraction based on Pulse-Coupled neural network. In Proceedings of the 2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference, Chongqing, China, 20–22 August 2011; pp. 186–189.
30. Lu, N. Dark convolutional neural network for forest smoke detection and localization based on single image. *Soft Comput.* **2022**, *26*, 8647–8659. [[CrossRef](#)]
31. He, L.J.; Gong, X.L.; Zhang, S.R.; Wang, L.J.; Li, F. Efficient attention based deep fusion CNN for smoke detection in fog environment. *Neurocomputing* **2021**, *434*, 224–238. [[CrossRef](#)]
32. Li, T.T.; Zhu, H.W.; Hu, C.H.; Zhang, J.G. An attention-based prototypical network for forest fire smoke few-shot detection. *J. For. Res.* **2022**, *33*, 1493–1504. [[CrossRef](#)]
33. Hu, Y.W.; Zhan, J.L.; Zhou, G.X.; Chen, A.B.; Cai, W.W. Fast forest fire smoke detection using MVMNet. *Knowl.-Based Syst.* **2022**, *241*, 108219. [[CrossRef](#)]
34. Zhang, L.J.; Lu, C.; Xu, H.W.; Chen, A.B.; Li, L.J.; Zhou, G.X. MMFNet: Forest Fire Smoke Detection Using Multiscale Convergence Coordinated Pyramid Network with Mixed Attention and Fast-robust NMS. *IEEE Internet Things J.* **2023**, *2023*, 18168–18180. [[CrossRef](#)]
35. Al-Smadi, Y.; Alauthman, M.; Al-Qerem, A.; Aldweesh, A.; Quaddoura, R. Early Wildfire Smoke Detection Using Different YOLO Models. *Machines* **2023**, *11*, 246. [[CrossRef](#)]
36. Zheng, X.; Chen, F.; Lou, L.M.; Cheng, P.L.; Huang, Y. Real-Time Detection of Full-Scale Forest Fire Smoke Based on Deep Convolution Neural Network. *Remote Sens.* **2022**, *14*, 536. [[CrossRef](#)]
37. Li, J.Y.; Zhou, G.X.; Chen, A.B.; Wang, Y.F.; Jiang, J.W.; Hu, Y.H.; Lu, C. Adaptive linear feature-reuse network for rapid forest fire smoke detection model. *Ecol. Inform.* **2022**, *68*, 101584. [[CrossRef](#)]
38. Li, A.; Zhao, Y.Q.; Zheng, Z.X. Novel Recursive BiFPN Combining with Swin Transformer for Wildland Fire Smoke Detection. *Forests* **2022**, *13*, 2032. [[CrossRef](#)]
39. Qian, J.J.; Lin, J.; Bai, D.; Xu, R.J.; Lin, H.F. Omni-Dimensional Dynamic Convolution Meets Bottleneck Transformer: A Novel Improved High Accuracy Forest Fire Smoke Detection Model. *Forests* **2023**, *14*, 838. [[CrossRef](#)]
40. Yuan, C.; Liu, Z.; Zhang, Y. Fire detection using infrared images for UAV-based forest fire surveillance. In Proceedings of the 2017 International Conference on Unmanned Aircraft Systems (ICUAS), Miami, FL, USA, 13–16 June 2017; pp. 567–572.
41. Ya’acob, N.; Najib, M.S.M.; Tajudin, N.; Yusof, A.L.; Kassim, M. Image Processing Based Forest Fire Detection using Infrared Camera. *J. Phys. Conf. Ser.* **2021**, *1769*, 012014. [[CrossRef](#)]
42. Yuan, C.; Liu, Z.X.; Zhang, Y.M. Aerial Images-Based Forest Fire Detection for Firefighting Using Optical Remote Sensing Techniques and Unmanned Aerial Vehicles. *J. Intell. Robot. Syst.* **2017**, *88*, 635–654. [[CrossRef](#)]
43. Wahyono; Harjoko, A.; Dharmawan, A.; Adhinata, F.D.; Kosala, G.; Jo, K.H. Real-Time Forest Fire Detection Framework Based on Artificial Intelligence Using Color Probability Model and Motion Feature Analysis. *Fire* **2022**, *5*, 23. [[CrossRef](#)]
44. Yang, X.B.; Hua, Z.C.; Zhang, L.; Fan, X.J.; Zhang, F.Q.; Ye, Q.L.; Fu, L.Y. Preferred vector machine for forest fire detection. *Pattern Recognit.* **2023**, *143*, 109722. [[CrossRef](#)]
45. Emmy Prema, C.; Vinsley, S.S.; Suresh, S. Efficient Flame Detection Based on Static and Dynamic Texture Analysis in Forest Fire Detection. *Fire Technol.* **2018**, *54*, 255–288. [[CrossRef](#)]
46. Muhammad, K.; Ahmad, J.; Baik, S.W. Early fire detection using convolutional neural networks during surveillance for effective disaster management. *Neurocomputing* **2018**, *288*, 30–42. [[CrossRef](#)]
47. Liu, Z.C.; Zhang, K.; Wang, C.Y.; Huang, S.Y. Research on the identification method for the forest fire based on deep learning. *Optik* **2020**, *223*, 165491. [[CrossRef](#)]
48. Muhammad, K.; Ahmad, J.; Lv, Z.H.; Bellavista, P.; Yang, P.; Baik, S.W. Efficient Deep CNN-Based Fire Detection and Localization in Video Surveillance Applications. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, *49*, 1419–1434. [[CrossRef](#)]
49. Jiao, Z.T.; Zhang, Y.M.; Xin, J.; Mu, L.X.; Yi, Y.M.; Liu, H.; Liu, D. A Deep Learning Based Forest Fire Detection Approach Using UAV and YOLOv3. In Proceedings of the 2019 1st International Conference on Industrial Artificial Intelligence (IAI), Shenyang, China, 23–27 July 2019; pp. 1–5.
50. Xue, Z.Y.; Lin, H.F.; Wang, F. A Small Target Forest Fire Detection Model Based on YOLOv5 Improvement. *Forests* **2022**, *13*, 1332. [[CrossRef](#)]
51. Chen, G.; Zhou, H.; Li, Z.Y.; Gao, Y.C.; Bai, D.; Xu, R.J.; Lin, H.F. Multi-Scale Forest Fire Recognition Model Based on Improved YOLOv5s. *Forests* **2023**, *14*, 315. [[CrossRef](#)]
52. Lin, J.; Lin, H.F.; Wang, F. A Semi-Supervised Method for Real-Time Forest Fire Detection Algorithm Based on Adaptively Spatial Feature Fusion. *Forests* **2023**, *14*, 361. [[CrossRef](#)]
53. Xue, Q.L.; Lin, H.F.; Wang, F. FCDM: An Improved Forest Fire Classification and Detection Model Based on YOLOv5. *Forests* **2022**, *13*, 2129. [[CrossRef](#)]
54. Li, J.H.; Xu, R.J.; Liu, Y.F. An Improved Forest Fire and Smoke Detection Model Based on YOLOv5. *Forests* **2023**, *14*, 833. [[CrossRef](#)]
55. Lu, K.J.; Huang, J.W.; Li, J.H.; Zhou, J.S.; Chen, X.L.; Liu, Y.F. MTL-FFDET: A Multi-Task Learning-Based Model for Forest Fire Detection. *Forests* **2022**, *13*, 1448. [[CrossRef](#)]
56. Huang, J.; He, Z.; Guan, Y.; Zhang, H. Real-Time Forest Fire Detection by Ensemble Lightweight YOLOX-L and Defogging Method. *Sensors* **2023**, *23*, 1894. [[CrossRef](#)]

57. Hopkins, B.; O'Neill, L.; Afghah, F.; Razi, A.; Rowell, E.; Watts, A.; Fule, P.; Coen, J. FLAME 2: Fire Detection and Modeling: Aerial Multi-Spectral Image Dataset. IEEE Dataport 2022. Available online: <https://iee-dataport.org/open-access/flare-2-fire-detection-and-modeling-aerial-multi-spectral-image-dataset> (accessed on 6 September 2023).
58. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 June 2016; pp. 770–778.
59. Wang, Q.L.; Wu, B.G.; Zhu, P.F.; Li, P.H.; Zuo, W.M.; Hu, Q.H. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 13–19 June 2020; pp. 11531–11539.
60. Yang, L.X.; Zhang, R.Y.; Li, L.; Xie, X.H. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In Proceedings of the International Conference on Machine Learning, Shenzhen, China, 18–24 July 2021; pp. 11863–11874.
61. Hou, Q.B.; Zhou, D.Q.; Feng, J.S. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
62. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.M.; Dollar, P. Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
63. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
64. Ren, S.Q.; He, K.M.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)]
65. Redmon, J.; Farhadi, A.J. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
66. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y. Yolov4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
67. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
68. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding Yolo Series in 2021. *arXiv* **2021**, arXiv:2107.08430.
69. Reis, D.; Kupec, J.; Hong, J.; Daoudi, A. Real-Time Flying Object Detection with YOLOv8. *arXiv* **2023**, arXiv:2305.09972.
70. Duan, K.W.; Bai, S.; Xie, L.X.; Qi, H.G.; Huang, Q.M.; Tian, Q. CenterNet: Keypoint Triplets for Object Detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October 2019–2 November 2019; pp. 6569–6578.
71. 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](#)]

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.