

## Article

# A High-Precision Ensemble Model for Forest Fire Detection in Large and Small Targets

Jiachen Qian <sup>1</sup>, Di Bai <sup>2,\*</sup>, Wanguo Jiao <sup>1</sup> , Ling Jiang <sup>1</sup>, Renjie Xu <sup>3</sup>, Haifeng Lin <sup>1,\*</sup>  and Tian Wang <sup>1</sup>

<sup>1</sup> The College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China; wgjiao@njfu.edu.cn (W.J.); jiangling@njfu.edu.cn (L.J.); wangtian@njfu.edu.cn (T.W.)

<sup>2</sup> College of Information Management, Nanjing Agricultural University, Nanjing 210095, China

<sup>3</sup> Department of Computing and Software, McMaster University, Hamilton, ON L8S 4L8, Canada; xur86@mcmaster.ca

\* Correspondence: baidi000@njau.edu.cn (D.B.); haifeng.lin@njfu.edu.cn (H.L.); Tel.: +86-25-8542-7827 (H.L.)

**Abstract:** Forest fires are major forestry disasters that cause loss of forest resources, forest ecosystem safety, and personal injury. It is often difficult for current forest fire detection models to achieve high detection accuracy on both large and small targets at the same time. In addition, most of the existing forest fire detection models are single detection models, and using only a single model for fire detection in a complex forest environment has a high misclassification rate, and the accuracy rate needs to be improved. Aiming at the above problems, this paper designs two forest fire detection models (named WSB and WSS) and proposes an integrated learning-based forest fire detection model (named WSB\_WSS), which also obtains high accuracy in the detection of forest fires with large and small targets. In order to help the model predict the location and size of forest fire targets more accurately, a new edge loss function, Wise-Faster Intersection over Union (WFloU), is designed in this paper, which effectively improves the performance of the forest fire detection algorithm. The WSB model introduces the Simple-Attention-Module (SimAM) attention mechanism to make the image feature extraction more accurate and introduces the bi-directional connectivity and cross-layer feature fusion to enhance the information mobility and feature expression ability of the feature pyramid network. The WSS model introduces the Squeeze-and-Excitation Networks (SE) attention mechanism so that the model can pay more attention to the most informative forest fire features and suppress unimportant features, and proposes Spatial Pyramid Pooling-Fast Cross Stage Partial Networks (SPPFCSPC) to enable the network to extract features better and speed up the operation of the model. The experimental findings demonstrate that the WSB model outperforms other approaches in the context of identifying forest fires characterized by small-scale targets, achieving a commendable accuracy rate of 82.4%, while the WSS model obtains a higher accuracy of 92.8% in the identification of large target forest fires. Therefore, in this paper, a more efficient forest fire detection model, WSB\_WSS, is proposed by integrating the two models through the method of Weighted Boxes Fusion (WBF), and the accuracy of detecting forest fires characterized by small-scale targets attains 83.3%, while for forest fires with larger dimensions, the accuracy reaches an impressive 93.5%. This outcome effectively leverages the strengths inherent in both models, consequently achieving the dual objective of high-precision detection for both small and large target forest fires concurrently.

**Keywords:** forest fire; integrated learning; target detection



**Citation:** Qian, J.; Bai, D.; Jiao, W.; Jiang, L.; Xu, R.; Lin, H.; Wang, T. A High-Precision Ensemble Model for Forest Fire Detection in Large and Small Targets. *Forests* **2023**, *14*, 2089. <https://doi.org/10.3390/f14102089>

Academic Editors: Jia Wang, Weiheng Xu, Jincheng Liu, Zhichao Wang, Stelian Alexandru Borz and Cate Macinnis-Ng

Received: 31 August 2023

Revised: 25 September 2023

Accepted: 17 October 2023

Published: 18 October 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

As of 2022, the country has a forest area of 231 million hectares, with a forest cover of 24.02% [1]. In the past five years, there have been 7301 forest fires in China. The forest, known as the lungs of the earth, is of vital importance to the earth's ecology. It maintains ecological stability by purifying harmful gases and releasing the oxygen necessary for the survival of plants and animals. In addition, forests regulate temperature, keep soil erosion

at bay, and maintain the diversity of biological species. Forest fire is a worldwide forestry disaster and the first of the three major forest natural disasters. Once a forest fire occurs, it will cause huge natural environmental losses and economic losses. Forest fires are the main forestry disaster causing loss of forest resources, forest ecosystem safety, and personal injury in China. And forest fire prevention is also an important part of disaster prevention and mitigation work in China [2]. Therefore, it is of great importance to guarantee the safety of forest resources, ecological safety, and people's lives and property.

China has formed a four-level, three-dimensional forest fire monitoring level, including ground patrol, near-ground monitoring, aerial patrol, and satellite monitoring. Ground patrol fire prevention measures mainly rely on forest rangers and forest fire prevention professionals to inspect around the clock, so manual inspection has some defects, such as heavy tasks, poor objectivity, and low efficiency, among others. Near-ground monitoring includes watchtower detection and remote video detection, which requires watchmen and monitoring personnel to use observation equipment for monitoring and analysis [3]. It requires a high technical level of monitoring personnel and easily produces monitoring blind spots. Aerial patrol monitoring mainly uses airplanes and drones to carry out high-altitude patrols, mainly used in the sparsely populated primitive forests. However, the cost of patrolling is high and cannot be monitored continuously on a large scale for a long time [4]. Satellite early warning monitoring monitors the heat sources of objects on the ground through artificial satellites, but it will be disturbed by factors such as high temperature, bare ground, and strong reflectors, resulting in misjudgment and omission of fire points [5].

Zheng et al. improved the dynamic convolution neural network and designed a forest fire prediction model that can accurately identify and classify the forest fire risk under the condition of natural light [6]. Yang et al. used vector machines for forest fire detection [7]. S. Sudhakar proposed a UAV-based method for forest fire detection with improved recognition algorithms applicable to UAVs to reduce the false alarms of forest fires [8]. Yoojin Kang et al. presented a model for forest fire detection using geosynchronous satellite data that effectively reduces detection delays and false alarms [9]. Ding proposes a new flame recognition color space that reduces the false alarm rate of flame recognition [10]. The method introduced by Zhao et al. put forth a novel approach for detecting diminutive targets of forest fires, thereby augmenting the precision in discerning such minute instances of forest fires [11]. Nouredine Moussa et al. proposed a novel routing protocol that makes the use of wireless sensor networks to detect forest fires more energy-efficient and reliable [12]. Zhang et al. orchestrated the development of a wireless sensor network tailored for the purpose of detecting forest fires. This endeavor involved the amalgamation of satellite surveillance, airborne observation, and terrestrial surveillance modalities [13].

Amidst the swift advancement and widespread adoption of technological innovations such as computer vision, image processing, and artificial intelligence, forest fire identification technology based on video images has attracted a lot of attention. The trajectory of development within forest fire recognition technology reliant upon video imagery can be primarily categorized into two cardinal avenues of progress. One is the background segmentation of the image, and then the theoretical research algorithm is designed; the features are extracted manually and sent to the classifier to recognize the feature image [14]. The second is image recognition based on deep learning, which sends the image into a convolutional neural network and automatically extracts features to recognize the image [15–19]. In recent years, many researchers have applied deep learning to forest fire detection. Sathishkumar et al. proposed a method to detect forest fires using forgettable learning based on deep learning [20]. Chen et al. used an adaptive copy–paste data enhancement method and a dice loss method to improve the semantic segmentation of forest flames [21]. Thus, compared with traditional fire monitoring methods, forest fire monitoring methods based on image recognition have a wide range of applications, large monitoring areas, lower cost, and higher accuracy. Although these studies have made good progress, there is still room for improvement in the effect of the model. At present, forest

fire identification often uses only a single model for monitoring. However, this method has a high rate of misidentification of forest fires in complex environments, and the accuracy still needs to be improved.

At present, within the domain of forest fire detection, the attainment of high-precision detection for forest fires characterized by diminutive targets continues to pose a formidable challenge. This challenge has led numerous scholars to immerse themselves in endeavors aimed at mitigating the issue of overlooking and misidentifying forest fires with small-scale targets. Consequently, an array of models has been innovatively formulated with a specific focus on the detection of forest fires characterized by these smaller targets. However, it is essential to recognize that forest fires distinguished by smaller-scale targets as opposed to their larger-scale counterparts exhibit conspicuous dissimilarities in terms of both spatial dimensions and inherent characteristics [22]; large target forest fires tend to have a large flame area, and their features are relatively easy to be captured in images or remote sensing data, whereas small-target forest fires usually have a small flame area and their features tend to be weak. Detection models designed for small-target forest fires may perform well in capturing the details and features of small targets but may experience performance degradation when dealing with large targets. This is because the model may focus too much on the details of small targets and ignore the overall characteristics of large targets. Forest fire detection models designed for large target forest fires are again prone to the problem of missed detection of small target forest fires. To summarize, the current forest fire detection model still lacks a model that comprehensively considers targets of different scales, and it is difficult to achieve high-precision detection of both large and small targets at the same time. In addition, the current forest fire identification often uses a single model for detection; however, the present approach is beset by a notable proclivity for misclassifying forest fires within intricate woodland settings, signifying a substantial misjudgment rate. Consequently, there persists a palpable imperative to enhance the precision of the methodology. Therefore, this paper designs an integrated learning-based forest fire detection model. This model is meticulously crafted to facilitate adaptive learning and the extraction of forest fire manifestations. The primary contributions of this paper are outlined as follows:

(1) In this paper, the WSS model, which is good at large target forest fire detection, and the WSB model, which is good at small target forest fire detection, are designed, and the two models are integrated based on the method of WBF, and a more efficient forest fire detection model, WSB\_WSS, is proposed. The integrated model gives full play to the advantages of the two models and is more robust and more generalized than the recognition results of a single model.

(2) A multi-scale forest fire detection method is designed to solve the problem of obvious differences in the scales of large and small targets. The information transfer and feature expression ability of the network is enhanced by two-way connection and cross-layer feature fusion. The introduction of SPPFCSPC allows the model to acquire spatial information at different scales, thus improving the perception of targets of different sizes.

(3) In order to optimize the model's anti-interference ability against noise, occlusion, and scale changes, a new edge loss function, WFloU Loss, is designed in this paper, which not only improves the training speed of the model but also improves the accuracy of forest fire detection.

(4) In order to amplify the model's emphasis on forest fires, this study introduces both the SimAM attention mechanism and the SE attention mechanism, which suppresses unimportant features and makes the model more accurately capture features related to forest fires.

## 2. Materials

### 2.1. Data Set

After investigation, it was found that there are no influential public datasets in the field of forest fires, and Yu et al. used drones to collect photographs of forest fires [23]. The

dataset employed in this study is sourced from the Fire Science Laboratory of the University of Science and Technology of China and the Fire Research Laboratory of Bilkent University. Moreover, the images obtained 98 randomly intercepting the forest fire video captured by the monitoring equipment, as well as a large number of network pictures and forest fire news pictures. The typical forest fire images in the dataset used in this paper are shown in Figure 1.



**Figure 1.** Example diagram of forest fire experimental dataset. (a,b) typical surface fire images, (c,d) Typical canopy fire images, (e–g) Forest fire images with both surface fire and canopy fire.

## 2.2. Data Annotation

How to label the data set directly affects the effect of model training and the accuracy of target detection. Three different labeling strategies were proposed by Lee et al. [24]. In this paper, taking forest fire as an example, as shown in Figure 2, the three labeling strategies are shown: (a) small area labeling method (labeling forest fires in pieces); (b) large area labeling method (labeling forest fires as a whole); and (c) hybrid labeling

method (labeling both large and small areas). In this paper, we use labeling software to manually label the forest fire images using the hybrid labeling method.



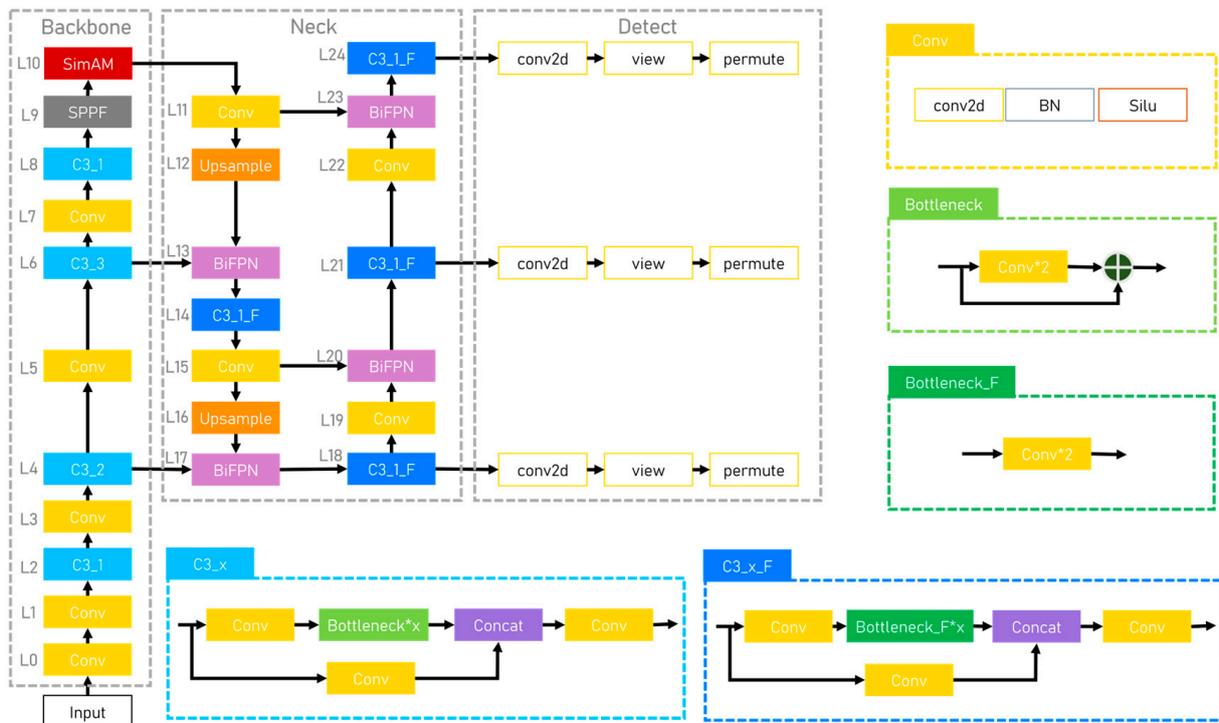
**Figure 2.** Example of annotation strategy for forest fire experimental data. The green boxes are the human-labeled object detection recognition boxes. (a) labeling forest fire areas in pieces, (b) labeling of forest fire areas as a whole, and (c) mixed labeling of large and small forest fire areas.

### 3. Methods

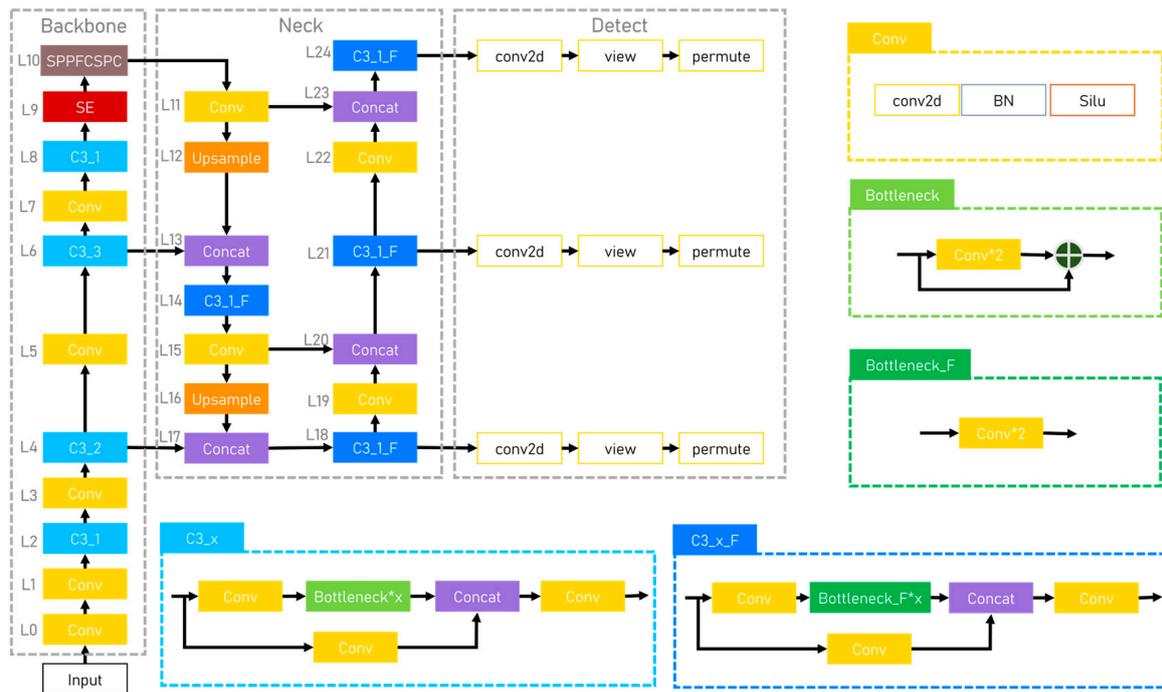
#### 3.1. The Models Proposed in This Paper—WSB and WSS

In this paper, two forest fire detection models are designed: WSB and WSS.

Both models are composed of four parts: Input, Backbone, Neck, and Detect. The Input and Detect parts of WSB and WSS are exactly the same, and there are slight differences in the Backbone and Neck. The network structure diagrams of the two models are shown in Figures 3 and 4.



**Figure 3.** Network structure diagram of the WSB model.



**Figure 4.** Network structure diagram of the WSS model.

The Input module mainly performs adaptive scaling processing, Mosaic data augmentation, and the computation of adaptive anchor frames applied to the images.

#### (1) Adaptive image scaling

In traditional target detection methods, images are usually scaled to a fixed size, which may result in the target being compressed or stretched, affecting the accuracy of detection. To solve this problem, this paper performs adaptive scaling operations based on the actual size of the target. Specifically, for each training sample, the input scales the image to an appropriate size based on the maximum edge length of the target in it. This can keep the shape of forest fire targets unchanged and improve the accuracy of forest fire target detection.

#### (2) Mosaic Data Enhancement

After adaptive image scaling, for each training sample, Mosaic data enhancement randomly selects four images and splices them together according to certain rules to form a large image. At the same time, according to the position information of the splicing, the image is adjusted with the corresponding label.

#### (3) Adaptive anchor frame calculation

In the target detection task, the anchor box is used to define the position and size of the target. Traditional target detection methods usually need to set up a set of anchor boxes with fixed size and scale in advance. In this paper, we adopt the adaptive anchor box calculation method, which clusters the target boxes in the training set by clustering algorithm to get a set of optimal anchor box sizes. This can make the anchor frames more closely match the true size distribution of the target, thus improving the accuracy of target detection.

The Backbone is the most important part of the network, which is responsible for extracting high-level semantic features from the input image. The first eight layers of the Backbone module for both the WSB and WSS models are layered with a series of C3\_X modules and convolutional modules, with only the last two layers differing. The penultimate layer of the WSB model is the Spatial Pyramid Pooling-Fast (SPPF), and the SimAM attention mechanism (Section 3.2) is added to the last layer of the Backbone network, which makes the extraction of image features more accurate; WSS adds SE (Section 3.3) to the penultimate layer of the Backbone network, which allows the model to pay more

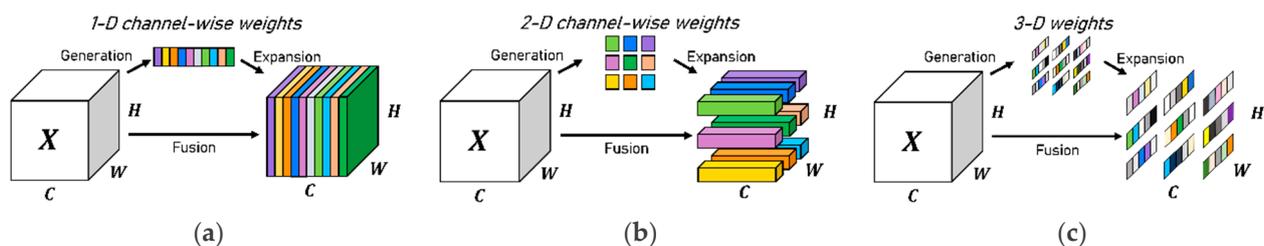
attention to the most informative forest fire features and suppresses unimportant features; and SPPFCSPC (Section 3.4) is used as the last layer of Backbone, which makes the network to be able to extract forest fire features more efficiently.

The Neck part of the WSS model consists of a series of convolution, sampling, C3\_X\_F module, and Concat module layers. Firstly, through the up-sampling operation, the feature information extracted by the network at the higher level is fused with the features at the lower level in a top-down manner. Then, the features at the lower level are fused with the feature information at the higher level in a bottom-up manner, which makes full use of semantic information at the higher level as well as the detailed information at the lower level for better feature fusion of the forest fire target. The Neck part of the WSS model, on the basis of the WSS model, has changed the Concat to a weighted bidirectional feature pyramid network structure Bi-directional Feature Pyramid Network (BiFPN) (Section 3.5), which assigns different weights to each input according to the importance of the features to achieve more effective feature fusion.

The WSS and WSB models have the same Detect module. The input comes from three feature maps extracted from layers 18, 21, and 24, which have different resolutions and semantic information, where the lower layer feature maps have higher spatial resolution and the higher layer feature maps have richer semantic information. For each feature map, the model uses Anchor Boxes to generate a set of candidate boxes. For each Anchor Box, the model calculates the confidence that it contains the target based on the predictions on the feature map. It also predicts the bounding box offsets of the target and the class probability of the target. These predictions are computed and adjusted by a set of convolutional and fully connected layers. In order to remove redundant detection results, this paper uses a Non-Maximum Suppression (NMS) algorithm to decide which boxes should be retained by traversing the candidate boxes and eliminating them when the overlap between a candidate box and an already retained candidate box exceeds a set threshold. Finally, for the retained candidate frames, the model selects the most likely category based on the category probability and adjusts the position and size of the candidate frames based on the prediction results of the bounding box loss function WFIoU Loss (Section 3.6) designed in this paper to more accurately localize forest fire targets.

### 3.2. SimAM Attention Mechanism

SimAM is a simple and parameter-free attention Module [25]. Different from a large number of scholars' research on feature extraction methods for spatial dimension or channel dimension, the SimAM module was devised to establish an energy function, thereby facilitating the derivation of three-dimensional attention weights without the need for additional parameters as feature maps, which ensures that it is lighter and more efficient. Figure 5 shows the step-by-step diagram of the different attentions. (a) The attention module generates 1-D weights by feature X, which is extended to channel attention. It differentiates treatment across distinct channels while uniformly considering all spatial dimensions. (b) The attention module generates 2-D weights by feature X, which is extended to spatial attention. It treats different spaces differently and treats all channels equally. (c) It is SimAM that generates 3-D weights.



**Figure 5.** Step-by-step diagram of different attentions. (a) Channel-wise Attention, (b) Spatial-wise Attention, (c) 3-D Weights for Attention.

SimAM finds important neurons by measuring the linear differentiability between neurons and assigns higher priority to these neurons. The equation defining the energy function of each neuron is shown in Equation (1).

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

In Equation (1):

$$\hat{t} = \omega_t t + b_t \tag{2}$$

$$\hat{x}_i = \omega_t x_i + b_t \tag{3}$$

$$M = H \times W \tag{4}$$

$\hat{t}$  is the target neuron in the channel,  $\hat{x}_i$  is the other neurons in the channel,  $i$  represents the index denoting the spatial dimension, and  $M$  signifies the count of neurons present within that specific channel. By binary labeling and adding the canonical term, the ultimate formulation of the energy function is depicted as presented in Equation (5).

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (w_t x_i + b_t))^2 + (1 - (w_t t + b_t))^2 + \lambda w_t^2 \tag{5}$$

By operating on individual neurons and integrating this linear separability into an end-to-end framework, a better enhancement to neural networks is achieved.

### 3.3. SE Attention Mechanism

At the heart of convolutional neural networks lies the fundamental convolution operation, which fuses features in spatial and channel dimensions by means of local perceptual fields. While feature extraction methods for the spatial dimension have been studied by a large number of scholars previously, SE is a channel attention mechanism that has been introduced to convolutional neural networks earlier [26]. The structure of the SE building block is depicted in Figure 6. SE is composed of Squeeze (Global Information Embedding) and Excitation (Adaptative Recalibration), respectively.

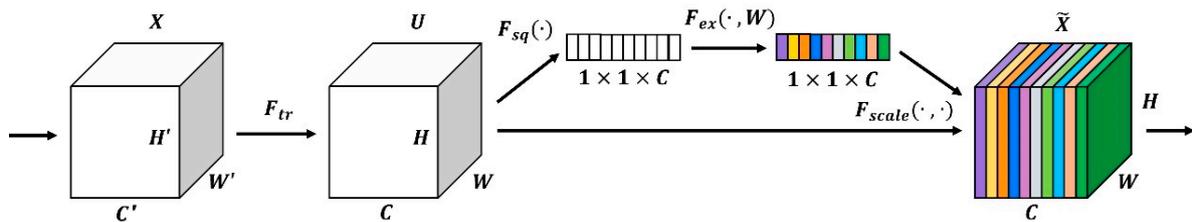


Figure 6. Structure of SE attention mechanism.

In order to make each cell of the transformed output U use contextual information outside this region, “Squeeze” employs global averaging pooling, which is the simplest aggregation technique that compresses the spatial dimensions H and W of U to 1 × 1. That is, the holistic spatial information is compacted into a channel descriptor through a global compression process, and one pixel is used to represent a channel to achieve low-dimensional embedding. The compressed feature is essentially a vector with only a channel dimension and no spatial dimension.

Squeeze is calculated as shown in Equation (6):

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \tag{6}$$

A subsequent operation is executed, endeavoring to comprehensively encapsulate interdependencies among channels to harness the insights extracted during the aggregation process in the squeeze operation. This endeavor necessitated the function to meet two essential criteria. Firstly, it required flexibility, particularly the capability to acquire knowledge of nonlinear interactions between channels. Secondly, it was imperative for the function to acquire a non-mutually exclusive association, as the objective was to facilitate the accentuation of multiple channels rather than enforcing exclusive activation of a solitary channel.

The Excitation component is realized by means of two fully connected layers. The initial fully connected layer condenses 'C' channels into 'C/r' channels, thereby mitigating computational demands (succeeded by ReLU activation). The subsequent fully connected layer restores the channel count to 'C' (succeeded by Sigmoid activation). R refers to the ratio of compression, also known as the dimensionality reduction rate. The Excitation operations are designed to fully capture the channel dependencies. For this purpose, a gating mechanism is designed with the following Equation (7).

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2\delta(W_1z)) \quad (7)$$

This function is able to learn nonlinear interactions between channels flexibly, as well as non-mutually exclusive relationships, which ensures that multiple channels are emphasized.

### 3.4. SPPFCSPC

The SPP structure, also known as spatial pyramid pooling, can convert arbitrarily sized feature maps into fixed-size feature vectors to achieve an adaptive size output [27]. In YOLOv5-6.0, the SPPF module is employed in lieu of the conventional SPP module. The SPPF module employs a sequence of cascading small-size pooling kernels as opposed to a singular large-size pooling kernel in the traditional SPP module. It not only retains the original function but also integrates the characteristic maps of different receptive fields. And in the case of enriching the expression ability of the feature map, the running speed is improved. The SPPCSPC structure is used in YOLOv7, which performs better than SPPF, but the number of parameters and computation is much higher, and the running speed is lower than SPPF. Figures 7 and 8 show the network structure diagrams of SPPF and SPPCSPC, respectively.

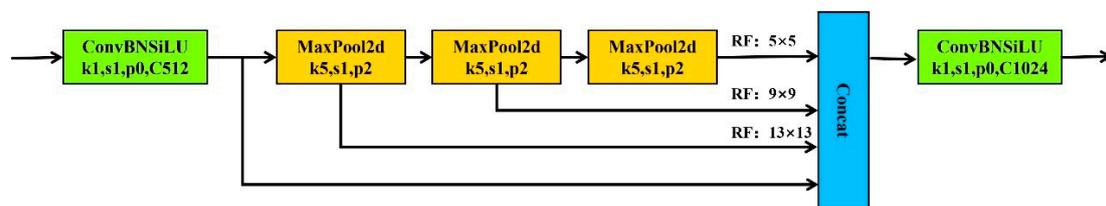


Figure 7. Network structure diagrams of SPPF.

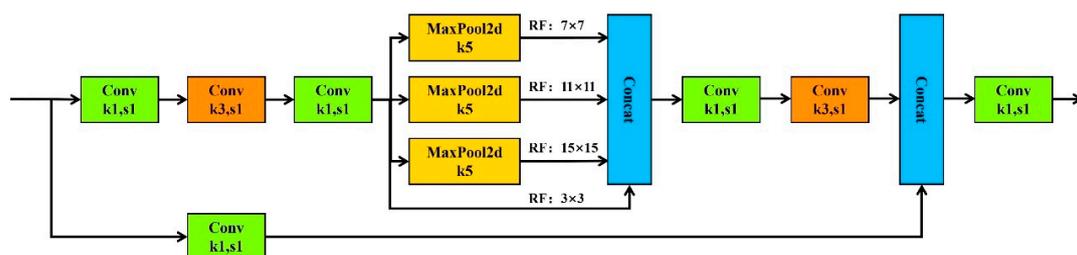


Figure 8. Network structure diagrams of SPPCSPC.

In this paper, the SPPCSPC is optimized by learning the idea of SPPF to obtain the SPPFCSPC, which improves the model running speed while keeping the perceptual field unchanged. Figure 9 shows the network structure of SPPFCSPC.

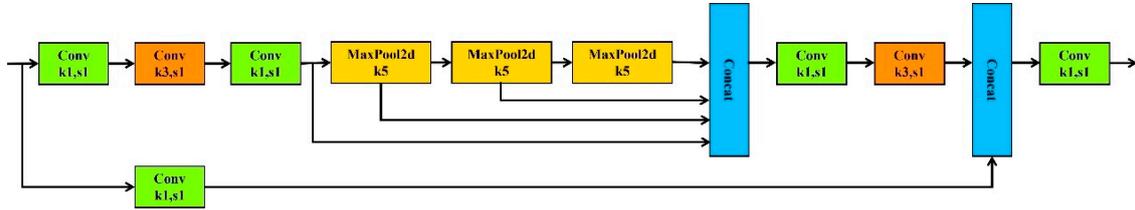


Figure 9. Network structure diagrams of SPPFCSPC.

3.5. BiFPN

BiFPN, known as Bidirectional Feature Pyramid Network, is a combination of top-down and bottom-up feature pyramid networks. Most of the models use the Feature Pyramid Network (FPN) plus the Path Aggregation Network (PAN) structure for feature fusion in the Neck layer fusing shallow and deep networks, while BiFPN can obtain a more effective feature fusion method compared to PAN+FPN, which can increase the coupling of features at each scale and perform better in shallow features for small target detection [28]. To address the issue of imprecise recognition stemming from the convergence of detection targets, the BiFPN implements a cross-scale connection methodology. This approach facilitates the modulation of feature representation for distinct detection features by means of cross-scale weights, thereby enabling both suppression and enhancement of feature expression. The structure diagrams of the FPN network, PAFPN network, and BiFPN network are shown in Figure 10.

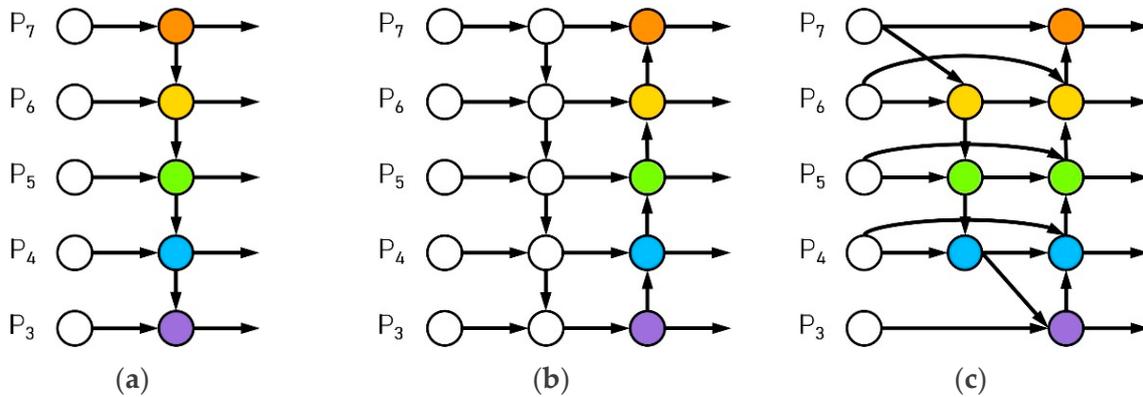


Figure 10. Comparison of three characteristic pyramidal network structures. (a) FPN, (b) PANet, (c) BiFPN.

The BiFPN structure is based on PAN, which deletes the nodes with no feature fusion and less contribution and adds a new channel between the original input node and the output node, thus saving resource consumption and fusing more feature information at the same time. The BiFPN structure carries on the fast normalized fusion through the ratio of the weight and the sum of the weight and finally normalizes the weight to [0, 1], which improves the perception ability of the target in different situations. At the same time, the information of feature maps between different levels can be fused at the prediction side, which can effectively solve the interference caused by noisy images and other factors.

3.6. Loss Function—WFIoU Loss

The loss function plays a crucial role in deep learning-based target detection. Through the process of minimizing the loss function, the model achieves expedited convergence,

leading to a reduction in the error rate of model predictions. Therefore, using a better loss function will better improve the detection performance of the model. This paper designs the loss function WFIoU based on Wise-IoU, which has a dynamic non-monotonic focusing mechanism. This mechanism has a smarter gradient gain allocation strategy, and this strategy can reduce the harmful gradients generated by low-quality examples. Through experiments, the performance of bounding box regression (BBR) loss without FMs is compared. From Figure 11 below, it can be observed that the WFIoU converges the fastest among a series of losses.

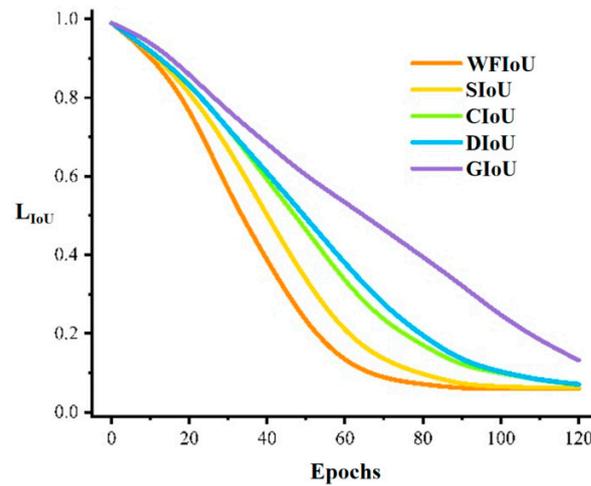


Figure 11.  $\mathcal{L}_{IoU}$  curves of BBR losses without FMs in simulation experiments.

WFIoUv1 constructs an attention-based bounding box loss, while v2 and v3 build on this by constructing a gradient gain (focus factor) calculation to attach a focus mechanism. WFIoUv1 is defined as shown in Equations (8) and (9):

$$\mathcal{L}_{WFIoUv1} = \mathcal{R}_{WFIoU} \mathcal{L}_{IoU} \tag{8}$$

$$\mathcal{R}_{WFIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \tag{9}$$

$W_g$  and  $H_g$  represent the width and height of the minimum enclosing frame, respectively. The  $\mathcal{L}_{IoU}$  of the ordinary quality anchor frame will be magnified by  $\mathcal{R}_{WFIoU}$ , while the  $\mathcal{R}_{WFIoU}$  of the high-quality anchor frame will be significantly reduced, separating  $W_g$  and  $H_g$  (operation with \* in the formula), which can prevent the gradient convergence from being hindered by  $\mathcal{R}_{WFIoU}$ .

In WFIoU v2, construct a monotonic focusing coefficient  $\mathcal{L}_{IoU}^{\gamma*}$  for  $\mathcal{L}_{WFIoUv2}$ . Define as Equation (10) below.

$$\mathcal{L}_{WFIoUv2} = \mathcal{L}_{IoU}^{\gamma*} \mathcal{L}_{WFIoUv1}, \gamma > 0 \tag{10}$$

In order to prevent the reduction of the convergence speed in the late training period, as shown in Equation (11), the mean value of  $\mathcal{L}_{IoU}$  is introduced as a normalization factor.

$$\mathcal{L}_{WFIoUv2} = \left(\frac{\mathcal{L}_{IoU}^*}{\mathcal{L}_{IoU}}\right)^\gamma \mathcal{L}_{WFIoUv1} \tag{11}$$

In order to prevent the low-quality example from producing a large harmful gradient, a smaller gradient gain is assigned to anchor frames with larger outliers in WFIoUv3. As shown in Equation (12), a non-monotonic focus factor based on  $\beta$  is constructed on the basis of v1.

$$\mathcal{L}_{WFIoUv3} = \alpha^{\phi(\beta) - \phi(\delta)} \mathcal{L}_{WFIoUv1} \tag{12}$$

In Equation (13):

$$\phi(x) = \log_a x - x \quad (13)$$

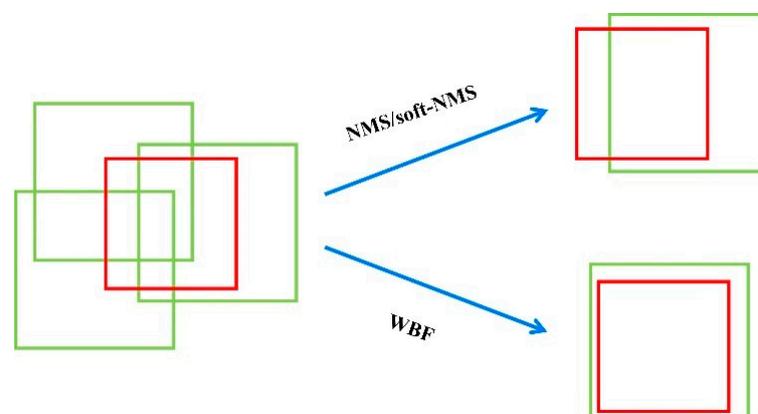
Because  $\mathcal{L}_{IoU}$  and the criteria for dividing anchor frames are not immutable, WFIoUv3 can change the gradient gain allocation strategy in a very timely manner to adapt to the current situation.

### 3.7. Forest Fire Detection Model WSB\_WSS Built on Integrated Learning

Integration learning is a commonly used method to improve the accuracy of model recognition [29]. The basic idea is to combine multiple base learners according to some method to get a strong learner. Generally, machine learning is used to get the most robust data model possible by selecting a suitable single model algorithm, but the single model enhancement has some limitations. Therefore, integration learning arises at the historic moment, which is used to fuse the results of multiple weak learners according to certain rules to further improve the accuracy of model recognition.

In this paper, two weakly supervised models, WSB and WSS, are designed. Through experiments, it is observed that WSB is more advantageous in the identification of small target forest fires, but the identification of large target forest fires is incomplete. WSS can identify large target forest fires in a complete way, but the detection accuracy of small target forest fires is not as good as that of the WSB model. Therefore, this paper integrates WSB and WSS and proposes a more efficient forest fire detection model, WSB\_WSS. The integrated model gives full play to the advantages of the two models, which can not only identify large forest fires completely but also accurately identify small target forest fires, which offsets the error between a single hypothesis and the target hypothesis and is more robust and generalizable than the recognition results of a single model.

The integrated model in this paper needs to fuse two different models. In the process of fusion, a large number of overlapping frames need to be de-duplicated to obtain the optimal detection frames. The commonly used de-duplication method is generally the non-maximum suppression method (NMS), but using NMS will directly eliminate the detection boxes with high overlap, resulting in some useful detection boxes being deleted. Therefore, in this paper, the Weighted Boxes Fusion (WBF) method is used to weight the fusion of the detection boxes, and then a new detection box is obtained to improve the accuracy of the final prediction box. As shown in Figure 12 below, the red box is the ground truth, and the green detection box is the prediction box. It can be seen that using NMS leaves only one box with the highest confidence but is not accurate while using WBF recalculates the confidence of the three prediction boxes to generate a new box, which is more accurate than NMS.



**Figure 12.** The model integrated method.

## 4. Results and Discussion

### 4.1. Experimental Environment and Parameter Setting

In this paper, the experimental environment is Windows 11 operating system, the CPU processor is AMD Ryzen 7 5800H, the GPU is NVIDIA GeForce RTX 3050, and the model is built and trained using Python 3.9, PyTorch 1.12.1, and CUDA 11.3. The resolution of the training and testing images for all experiments was 640 pixels  $\times$  640 pixels. The training process is 300 epochs, and the Initial learning rate is set to 0.01. The *IOU* threshold is set to 0.5 in the WBF algorithm, and the weight value is set to 1 for both models, which means that the results of both models have the same weight.

### 4.2. Evaluation Index

The commonly used evaluation indexes in neural networks are P (accuracy), R (recall rate), AP (average precision), and mAP (mean average precision). The calculation formula is shown below.

$$P = \frac{TP}{FP + TP} \quad (14)$$

$$R = \frac{TP}{TP + FN} \quad (15)$$

$$AP = \int_0^1 P(t) dt \quad (16)$$

$$mAP = \frac{1}{m} \sum_{t=1}^m AP(t) \quad (17)$$

In the formula, TP signifies the count of instances that are accurately classified as positive examples, which means the number of fires labeled as forest fires and correctly identified as forest fires by the model. TN signifies the count of instances that are correctly identified as negative cases, which means the number of cases not actually labeled as forest fire and successfully identified as not forest fire by the model. FP signifies the count of instances that are mistakenly identified as positive cases, which means the number of cases that are not actually labeled as forest fires but are incorrectly identified as such by the model. FN signifies the count of cases that are mistakenly identified as negative, which means the number of cases that are actually forest fires but are incorrectly identified as not forest fires by the model.

### 4.3. Experimental Comparison

This paper carried out ablation comparison experiments under the homemade forest fire dataset to validate the efficacy of the algorithm formulated within this research paper for the purpose of forest fire identification, and the results of the experiments are shown in Table 1. Among them, according to the scale requirements under the target size classification of the evaluation index of the COCO dataset, we define the target area as less than or equal to 32<sup>2</sup> as small target forest fires and define the target area as greater than or equal to 96<sup>2</sup> as large target forest fires. The results of the ablation experiments show that the SE mechanism is essential for improving the detection performance of large target fires in the model; the SimAM attention mechanism helps to improve the detection performance of large target fires in the model; and the SimAM attention mechanism helps to improve the detection performance of large target fires in the model. The SimAM attentional mechanism helps to improve the detection of small target forest fires; the introduction of the WFIoU loss function improves the detection accuracy of the model for both large target forest fire and small target forest fires; the WSS model incorporating the SE mechanism, BiFPN, and WFIoU loss function is better at detecting large target forest fires; the WSS model incorporating the SimAM Attention mechanism, SPPFCSPC module, and WFIoU loss function, the WSB model is better than the WSS model in detecting small target forest fires;

the WSB\_WSS model based on integrated learning proposed in this paper combines the advantages of the WSS model and the WSB model, which makes the detection accuracy of forest fire targets at different scales higher than that of the pre-integration model. The integrated model achieves 83.3% detection accuracy for small target forest fires and 93.5% detection accuracy for large target forest fires, which is 3.1% and 4.2% higher than the detection accuracy of the network model without the addition of each module, respectively.

**Table 1.** Results of ablation experiments.

	SimAM	SE	SPPFCSPC	BiFPN	WFIoU	mAP		
						Forest Fire	Small Target Forest Fire	Large Target Forest Fire
1	×	×	×	×	×	0.847	0.802	0.893
2	×	✓	×	×	×	0.841	0.762	0.919
3	✓	×	×	×	×	0.849	0.814	0.883
4	×	×	×	×	✓	0.855	0.811	0.898
5	×	✓	×	✓	✓	0.853	0.778	0.928
6	✓	×	✓	×	✓	0.863	0.824	0.901
7	✓	✓	✓	✓	✓	0.884	0.833	0.935

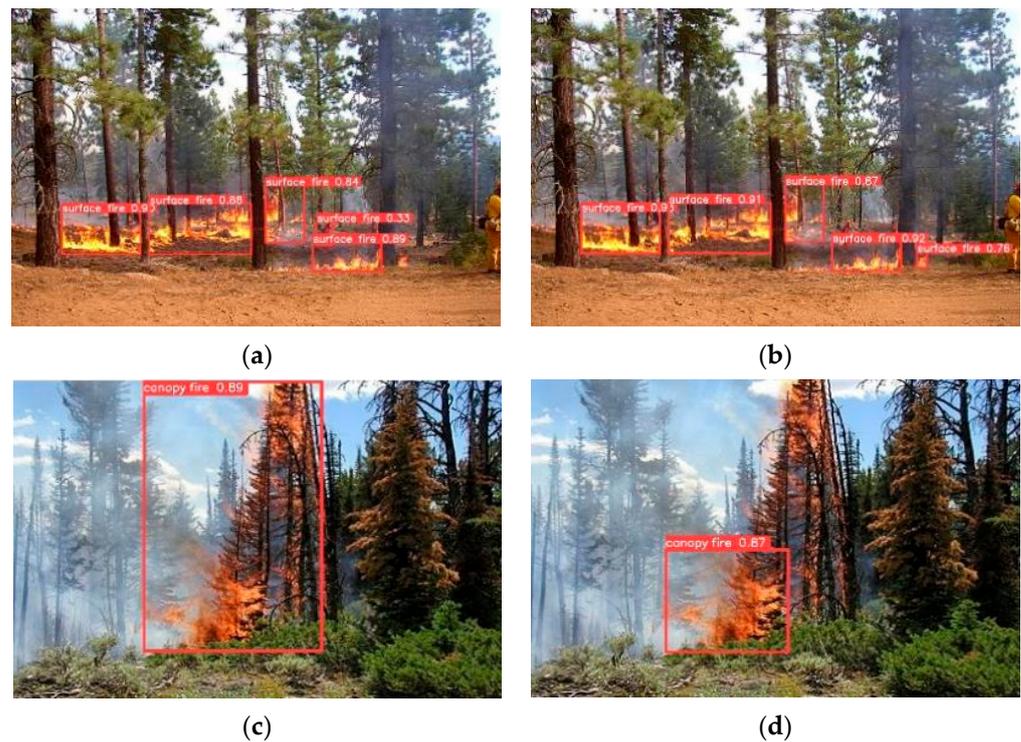
In order to objectively and comprehensively evaluate the performance of the WSB\_WSS model established in this paper in the forest fire detection task, this paper carries out comparative experiments with mainstream target detection models YOLOv3, YOLOv4, YOLOv5, and EfficientDet, and the results of the comparative experiments are shown in Table 2. From the results of the comparative experiments, it can be seen that the WSB\_WSS model established in this paper has a significantly higher forest fire detection accuracy than the mainstream models, and the accuracy of large and small target forest fire detection is also better than the mainstream models. Detection accuracy is significantly higher than that of the mainstream models, and the accuracy of large and small target forest fire detection is also better than that of the mainstream models. In addition, the integrated WSB\_WSS model combines the advantages of the WSB model and the WSS model, which not only has a high forest fire detection accuracy but also is better at detecting large and small target forest fires.

**Table 2.** Comparative test results with mainstream models.

Models	mAP		
	Forest Fire	Small Target Forest Fire	Large Target Forest Fire
YOLOv3	0.748	0.716	0.781
YOLOv4	0.796	0.776	0.817
YOLOv5	0.847	0.802	0.893
EfficientDet	0.851	0.813	0.889
WSB	0.853	0.778	0.928
WSS	0.863	0.824	0.901
WSB_WSS	0.884	0.833	0.935

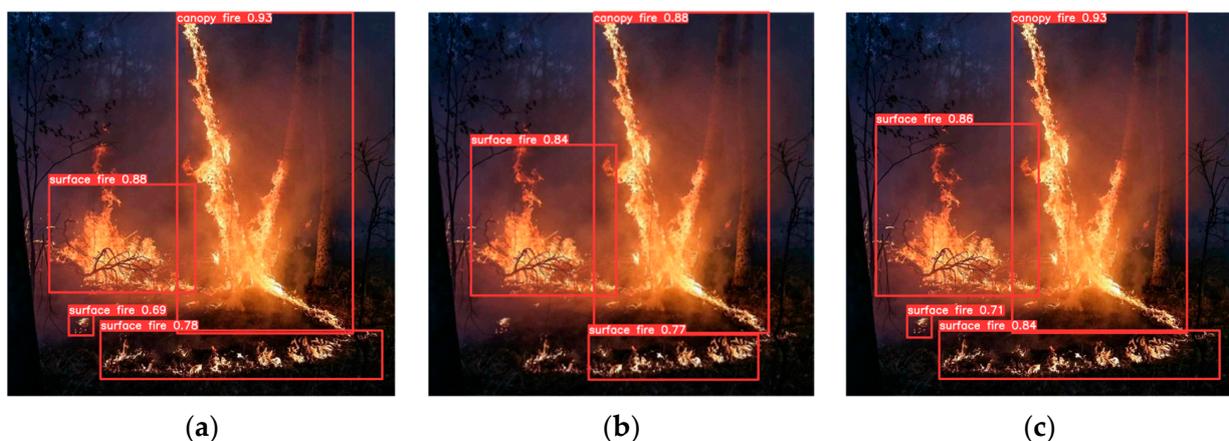
#### 4.4. Comparison of Recognition Results

This paper first compares the forest fire recognition effect of the WSS and WSB models. The WSS model not only misses a small target surface fire but also has a small target surface fire misdetection. In contrast, the WSB model has a better detection effect for small target forest fires and does not have the situation of miss detection and misdetection. When comparing Figure 13c,d, it can be seen that the WSS model can identify the large target trunk fire completely, while the WSB model only identifies part of the large target trunk fire. Based on the above recognition results, this paper considers integrating and fusing the WSS and WSB models to complement each other's strengths and obtain a model that is good at detecting large-area forest fires and does not miss detecting small target forest fires.



**Figure 13.** Comparison of recognition results between WSS and WSB. (a,c) show the detection results of the WSS model, while (b,d) show the detection results of the WSB model.

In order to feel the effectiveness of integrated learning more intuitively, this paper uses the same picture to compare the forest fire recognition effects of WSB, WSS, and the integrated model WSB\_WSS. Comparative analysis shows that the WSB model can accurately identify small target surface fires, but the identification of large target forest fires is sometimes incomplete, as shown in Figure 14a; the WSS model can completely identify large surface fires and trunk fires but misses part of the small target surface fires, as shown in Figure 14b; the integrated model can not only completely identify large forest fires, but also accurately identify surface fires and small target forest fires, as shown in Figure 14c; the integrated model can not only completely identify large forest fires, but also accurately identify surface fires and small target forest fires, as shown in Figure 14d and small target forest fires, as shown in Figure 14c.



**Figure 14.** (a) shows the detection results of the WSB model; (b) shows the detection results of the WSS model; (c) shows the detection results of the integrated model WSB\_WSS.

## 5. Conclusions

At present, the establishment of a high-precision detection model that can realize both large and small target forest fires is still a major difficulty in the field of forest fire detection. There are obvious differences in the scales and characteristics of large and small target fires, and the existing forest fire detection models still have the problem of not being able to comprehensively consider the targets at different scales. Detection models designed for small target forest fires often pay too much attention to the details of small targets, thus ignoring the overall characteristics of large target forest fires; models designed for large target forest fires are often prone to the omission of small-target forest fires. In addition, most of the current forest fire detection uses a single model for detection, which cannot adequately handle the impact of target size change, occlusion, and other factors on forest fire detection, and the robustness of the model is poor.

In order to solve the problem of the obvious difference in the scales of large and small target forest fires, this paper designs a multi-scale forest fire detection method, which enhances the feature expression ability of the network by means of bidirectional links and cross-layer feature fusion. In addition, this paper introduces SPPFCSPC to improve the model's ability to perceive targets at different scales. In addition, this paper proposes a new edge loss function, WFIoU Loss, to optimize the model's anti-interference ability against noise, occlusion, and scale changes and to improve the accuracy of the model for forest fire detection. This paper also introduces the SimAM attention mechanism and SE attention mechanism to increase the model's attention to forest fires. Based on the above techniques, this paper proposes two forest fire detection models, WSB and WSS. Through experiments, we find that the WSB model can accurately identify small target surface fires, but the identification of large target forest fires is sometimes incomplete, while the WSS model can completely identify large surface fires and trunk fires but misses part of the small target surface fires. Therefore, this paper integrates the two models based on the WBF method and designs a more efficient forest fire detection model, WSB\_WSS, which simultaneously realizes high-accuracy large target forest fire detection and small target forest fire detection. The model WSB\_WSS designed in this paper has an accuracy of 83.3% for small target forest fire detection and 93.5% for large target forest fire detection, which is significantly better than the existing mainstream models.

Considering that there are many smoke interferences in the process of forest fire recognition, which affect the presentation of forest fire images and the recognition and extraction of forest fire image features, and influence the accuracy of recognition, further analysis and experiments on how to exclude the influence of smoke features on forest fire images will be carried out in the following to improve the generalizability of the model [30,31].

**Author Contributions:** J.Q. devised the programs, drafted the initial manuscript, and contributed to writing embellishments. T.W. and R.X. helped with data collection and data analysis. H.L. and D.B. designed the project and revised the manuscript. W.J. and L.J. assisted in building the framework of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Key Research and Development Plan of Jiangsu Province (Grant No. BE2021716) and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (2023).

**Data Availability Statement:** All data generated or presented in this study are available upon request from the corresponding author. Furthermore, the models and code used during the study cannot be shared at this time as the data also form part of an ongoing study.

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

## References

1. Office of the National Greening Committee. Bulletin of China's land greening status in 2022. *Land Green.* **2023**, *3*, 6–11.
2. Chen, J.; Di, X.Y. Forest fire prevention management legal regime between China and the United States. *J. For. Res.* **2015**, *26*, 447–455. [[CrossRef](#)]
3. Rego, F.C.; Catry, F.X. Modelling the effects of distance on the probability of fire detection from lookouts. *Int. J. Wildland Fire* **2006**, *15*, 197–202. [[CrossRef](#)]
4. Hossain, F.M.A.; Zhang, Y.; Yuan, C. A Survey on Forest Fire Monitoring Using Unmanned Aerial Vehicles. In Proceedings of the 3rd International Symposium on Autonomous Systems (ISAS), Shanghai, China, 29–31 May 2019; pp. 484–489.
5. Gure, M.; Ozel, M.E.; Yildirim, H.H.; Ozdemir, M. Use of Satellite Images for Forest Fires in Area Determination and Monitoring. In Proceedings of the 4th International Conference on Recent Advances in Space Technologies, Istanbul, Turkey, 11–13 June 2009; pp. 27–32.
6. Zheng, S.; Gao, P.; Wang, W.; Zou, X. A Highly Accurate Forest Fire Prediction Model Based on an Improved Dynamic Convolutional Neural Network. *Appl. Sci.* **2022**, *12*, 6721. [[CrossRef](#)]
7. Yang, X.; Hua, Z.; Zhang, L.; Fan, X.; Zhang, F.; Ye, Q.; Fu, L. Preferred Vector Machine for Forest Fire Detection. *Pattern Recognit.* **2023**, *143*, 109722. [[CrossRef](#)]
8. Sudhakar, S.; Vijayakumar, V.; Kumar, C.S.; Priya, V.; Ravi, L.; Subramaniaswamy, V. Unmanned Aerial Vehicle (UAV) based Forest Fire Detection and monitoring for reducing false alarms in forest-fires. *Comput. Commun.* **2020**, *149*, 1–16. [[CrossRef](#)]
9. Kang, Y.; Jang, E.; Im, J.; Kwon, C. A deep learning model using geostationary satellite data for forest fire detection with reduced detection latency. *Gisci. Remote Sens.* **2022**, *59*, 2019–2035. [[CrossRef](#)]
10. Ding, X.; Gao, J. A new intelligent fire color space approach for forest fire detection. *J. Intell. Fuzzy Syst.* **2022**, *42*, 5265–5281. [[CrossRef](#)]
11. Zhao, L.; Zhi, L.; Zhao, C.; Zheng, W. Fire-YOLO: A Small Target Object Detection Method for Fire Inspection. *Sustainability* **2022**, *14*, 4930. [[CrossRef](#)]
12. Moussa, N.; Nurellari, E.; El Alaoui, A.E.B. A novel energy-efficient and reliable ACO-based routing protocol for WSN-enabled forest fires detection. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *14*, 11639–11655. [[CrossRef](#)]
13. Zhang, J.; Li, W.; Yin, Z.; Liu, S.; Guo, X. Forest fire detection system based on wireless sensor network. In Proceedings of the 2009 4th IEEE Conference on Industrial Electronics and Applications, Xi'an, China, 25–27 May 2009; pp. 520–523.
14. Liu, D.; Li, Q. Research And Realization of Aluminum Plate Surface Defects Classification Based on a Combination of BP Neural Network and SVM. In Proceedings of the International Conference on Chemical, Material and Food Engineering (CMFE), Kunming, China, 25–26 July 2015; pp. 624–627.
15. Zhai, H. Research on Image Recognition Based on Deep Learning Technology. In Proceedings of the 4th International Conference on Advanced Materials and Information Technology Processing (AMITP), Guilin, China, 24–25 September 2016; pp. 266–270.
16. Yuan, N.; Kang, B.H.; Xu, S.; Yang, W.; Ji, R. Research on Image Target Detection and Recognition Based on Deep Learning. In Proceedings of the International Conference on Information Systems and Computer Aided Education (ICISCAE), Changchun, China, 6–8 July 2018; pp. 158–163.
17. Lin, H.; Qian, J.; Di, B. Learning for Adaptive Multi-Copy Relaying in Vehicular Delay Tolerant Network. *IEEE Trans. Intell. Transp. Syst.* **2023**, 1–10. [[CrossRef](#)]
18. Wang, Y.; Yang, X.; Zhang, L.; Fan, X.; Ye, Q.; Fu, L. Individual tree segmentation and tree-counting using supervised clustering. *Comput. Electron. Agric.* **2023**, *205*, 107629. [[CrossRef](#)]
19. Sumathi, D.; Alluri, K. Deploying deep learning models for various real-time applications using Keras. In *Advanced Deep Learning for Engineers and Scientists: A Practical Approach*; Springer: Cham, Switzerland, 2021; pp. 113–143.
20. 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](#)]
21. Chen, B.; Bai, D.; Lin, H.; Jiao, W. FlameTransNet: Advancing Forest Flame Segmentation with Fusion and Augmentation Techniques. *Forests* **2023**, *14*, 1887. [[CrossRef](#)]
22. Zhan, J.; Hu, Y.; Cai, W.; Zhou, G.; Li, L. PDAM-STPNet: A Small Target Detection Approach for Wildland Fire Smoke through Remote Sensing Images. *Symmetry* **2021**, *13*, 2260. [[CrossRef](#)]
23. Yu, C.; Chen, L. Forest Fire Detection Method Based on Uav Image Acquisition and Processing. *Fresenius Environ. Bull.* **2021**, *30*, 13166–13172.
24. Lee, Y.; Im, D.; Shim, J. Data Labeling Research for Deep Learning Based Fire Detection System. In Proceedings of the 4th International Conference on Systems of Collaboration Big Data, Internet of Things & Security (SysCoBioTS), Casablanca, Morocco, 12–13 December 2019; pp. 1–4.
25. Yang, L.X.; Zhang, R.Y.; Li, L.D.; Xie, X.H. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In Proceedings of the International Conference on Machine Learning (ICML), Virtual, 18–24 July 2021.
26. Jie, H.; Li, S.; Gang, S.; Albanie, S. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
27. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)] [[PubMed](#)]

28. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
29. Dietterich, T.G. Ensemble methods in machine learning. In *Multiple Classifier Systems*; Kittler, J., Roli, F., Eds.; Lecture Notes in Computer Science; Springer: Boston, MA, USA, 2000; Volume 1857, pp. 1–15.
30. Chen, G.; Cheng, R.; Lin, X.; Jiao, W.; Bai, D.; Lin, H. LMDFS: A Lightweight Model for Detecting Forest Fire Smoke in UAV Images Based on YOLOv7. *Remote Sens.* **2023**, *15*, 3790. [[CrossRef](#)]
31. Wang, X.; Li, M.; Gao, M.; Liu, Q.; Li, Z.; Kou, L. Early smoke and flame detection based on transformer. *J. Saf. Sci. Resil.* **2023**, *4*, 294–304. [[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.