


Article

# FCDM: An Improved Forest Fire Classification and Detection Model Based on YOLOv5

Qilin Xue <sup>1</sup>, Haifeng Lin <sup>1,\*</sup>  and Fang Wang <sup>2,\*</sup>

<sup>1</sup> College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China

<sup>2</sup> College of Electronic Engineering, Nanjing Xiaozhuang University, Nanjing 211171, China

\* Correspondence: haifeng.lin@njfu.edu.cn (H.L.); wangfang0182217@njzc.edu.cn (F.W.);  
Tel.: +86-25-8542-7827 (H.L.); +86-25-86175539 (F.W.)

**Abstract:** Intense, large-scale forest fires are damaging and very challenging to control. Locations, where various types of fire behavior occur, vary depending on environmental factors. According to the burning site of forest fires and the degree of damage, this paper considers the classification and identification of surface fires and canopy fires. Deep learning-based forest fire detection uses convolutional neural networks to automatically extract multidimensional features of forest fire images with high detection accuracy. To accurately identify different forest fire types in complex backgrounds, an improved forest fire classification and detection model (FCDM) based on YOLOv5 is presented in this paper, which uses image-based data. By changing the YOLOv5 bounding box loss function to Siou Loss and introducing directionality in the cost of the loss function to achieve faster convergence, the training and inference of the detection algorithm are greatly improved. The Convolutional Block Attention Module (CBAM) is introduced in the network to fuse channel attention and spatial attention to improve the classification recognition accuracy. The Path Aggregation Network (PANet) layer in the YOLOv5 algorithm is improved into a weighted Bi-directional Feature Pyramid Network (BiFPN) to fuse and filter forest fire features of different dimensions to improve the detection of different types of forest fires. The experimental results show that this improved forest fire classification and identification model outperforms the YOLOv5 algorithm in both detection performances. The mAP@0.5 of fire detection, surface fire detection, and canopy fire detection was improved by 3.9%, 4.0%, and 3.8%, respectively. Among them, the mAP@0.5 of surface fire reached 83.1%, and the canopy fire detection reached 90.6%. This indicates that the performance of our proposed improved model has been effectively improved and has some application prospects in forest fire classification and recognition.



**Citation:** Xue, Q.; Lin, H.; Wang, F. FCDM: An Improved Forest Fire Classification and Detection Model Based on YOLOv5. *Forests* **2022**, *13*, 2129. <https://doi.org/10.3390/f13122129>

Academic Editor: Michele Salis

Received: 3 November 2022

Accepted: 10 December 2022

Published: 12 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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/>).

**Keywords:** forest fire classification detection; YOLOv5; Siou Loss; CBAM; BiFPN

## 1. Introduction

Forests are the ecosystem with the largest terrestrial distribution area and the richest biological types. With the role of purifying water, regulating climate and enriching types of diversity, they are an important guarantee for maintaining the ecological security of the Earth. And substantial forest fires are the more terrible forest disasters and can result in significant property losses and severe safety issues [1].

Affected by environmental conditions, such as weather, topography, fuels, etc., locations where different types of fire behavior occur are different, and the corresponding fire suppression measures are different. A typical fire classification is based on the fire burning site, and the forest fire types can be classified as underground fires, surface fires and canopy fires. Among them, underground fires [2] mostly spread and expand in the eroded or peat layers below the ground surface. It is not easy to see the fire during daytime, and it is more difficult to detect based on computer vision recognition. Surface fire [3] is a relatively common type of forest fire that spreads mostly along the forest floor surface,

burning ground covers, such as young trees, shrubs, and understory, and burning the base of large tree trunks and exposed roots. When surface fires encounter strong winds or groups of young coniferous trees, dead-standing trees, and low-hanging branches, the flames burn to the upper part of the canopy and expand rapidly downwind to become canopy fires. Canopy fires are extremely devastating to the forest because they spread along the canopy, burning ground cover, young trees, and underwood below while destroying leaves, burnt branches, and trunks above. In this paper, forest fire images are based on computer vision images taken by UAVs with a small pixel scale, so surface fires and canopy fires are mainly considered.

Traditional forest fire monitoring methods include ground detection, aerial monitoring, satellite monitoring, etc. Ground monitoring carries out real-time monitoring of forest fires through manual inspection, radar monitoring [4], etc. The recognition rate is high, but it is susceptible to the influence of terrain and natural conditions, and the detection efficiency is low. Aerial monitoring [5–7] occurs through low-altitude aircraft such as drones carrying forest fire monitoring equipment to achieve real-time monitoring of forest fires on a large scale. However, the detection cost is high, vulnerable to environmental factors, and it is difficult to carry out long time continuous monitoring. Satellite remote sensing technology [8–10] covers a wide range of areas and has a short imaging period, which gradually becomes an important means of forest fire monitoring. However, it is limited by the repetition period and detection resolution, which makes it difficult to achieve continuous monitoring and is less flexible in mobility.

Currently, forest fire detection based on video images uses a large amount of existing video surveillance equipment to achieve 24 h monitoring of forest areas [11–13]. It can achieve large range observation, high detection accuracy, easy data storage, etc., which is a more reasonable and low-cost solution in forest fire detection technology. A forest fire detection tracking program was created by Chi Yuan et al. [14] using the chromaticity characteristics of flames, channels in the Lab color model, and a combination of median filtering and color space conversion. Jin et al. [15] were able to extract the fire. Combining multicolor detection in RGB, HIS, and YUV color spaces with motion features might help locate potential fire locations. Prema et al. [16] extracted static and dynamic texture features from candidate fire areas based on colors in YVbCr for various fire detection scenarios. MAI Mahmoud et al. [17] applied background subtraction to fire boundaries, and used a color segmentation model to label and classify regions into true fire and non-fire using support vector machines.

Video and image-based forest fire detection can be divided into traditional machine learning-based [18] and deep learning-based forest fire inspection. Due to the complex and versatile forest environments, deep learning-based forest fire inspection can automatically extract multidimensional forest fire features using deep convolutional neural networks, which have better classification, generalization, and detection capabilities than traditional machine learning methods. Chen et al. [19] firstly used a local binary pattern (LBP) feature extraction and support vector machine (SVM) classifier for smoke inspection and presented a method for forest fire inspection based on UAV images using a convolutional neural network (CNN). Panagiotis et al. [20] identified candidate fire areas by a Faster R-CNN network and proposed an image-based fire detection method combining deep learning networks and multidimensional texture analysis. PuLi et al. [21] proposed an image-based fire detection algorithm with a target detection CNN model based on existing algorithms such as Faster-RCNN, R-FCN, SSD, and YOLO V3. A custom framework was proposed by Saima Majid et al. [22] which uses Grad-CAM method for visualization and localization of fire images and uses attention mechanism and migration learning along with EfficientNetB0 network training. Barisic et al. [23] applied a convolutional neural network YOLO for training to propose a vision-based UAV real-time detection and tracking system. However, existing forest fire detection has fewer experiments to classify fires. Therefore, the deep learning-based forest fire classification and inspection technique helps to detect and identify

forest fire types as early as possible and take relevant suppression measures as soon as possible, which has important practical application value.

In this paper, we develop an improved forest fire classification and detection model based on YOLOv5, which uses image-based data. First, the YOLOv5 [24] bounding box loss function CIOU Loss [25] is changed to SIOU Loss [26]. Directionality is introduced in the cost of the loss function to achieve faster convergence, which significantly improves the training and inference of the detection algorithm. Secondly, due to the small target and low pixels of forest fires, which are prone to information loss, the CBAM [27] attention mechanism module is introduced in the network. The fusion of channel attention and spatial attention improves the accuracy of extracted features and enhances the classification recognition accuracy. Finally, the Path Aggregation Network (PANet) [28] layer is improved into a weighted Bi-directional Feature Pyramid Network (BiFPN) [29] to fuse and filter forest fire features of different dimensions to prevent feature loss and enhance the detection of different types of forest fires.

The rest of this paper is organized as follows. In Section 2, the homemade forest fire classification dataset and the methods and modules used in the experiments are introduced. The forest fire classification detection model presented in this paper is also elaborated in this section. Section 3 presents the metrics for evaluating the model performance. This section also includes the experimental results of each part of the improvement. Section 4 describes the discussion and analysis of the model, as well as the outlook for future work. A summary of the entire work is presented in Section 5.

## 2. Materials and Methods

### 2.1. Dataset

The quantity and quality of datasets largely affect the results of improved forest fire classification recognition based on YOLOv5. In this paper, forest fire images are based on computer vision images taken by drones. Therefore, we first find forest fire images through the network as well as obtain forest fire images from some publicly available datasets. The images suitable for training are manually filtered. After that, we perform manual classification. Surface fires start burning from ground cover and spread along the ground surface. Canopy fires burn fiercely and tend to be found within coniferous forests with high resinous content. The collected forest fire images are classified into surface fire images and canopy fire images depending on the burning location. Samples of typical forest fire images and forest fire images from the UAV view are shown in Figure 1.



(a) Typical surface fire



(b) Surface fire from the UAV view

Figure 1. Cont.



(c) Typical canopy fire



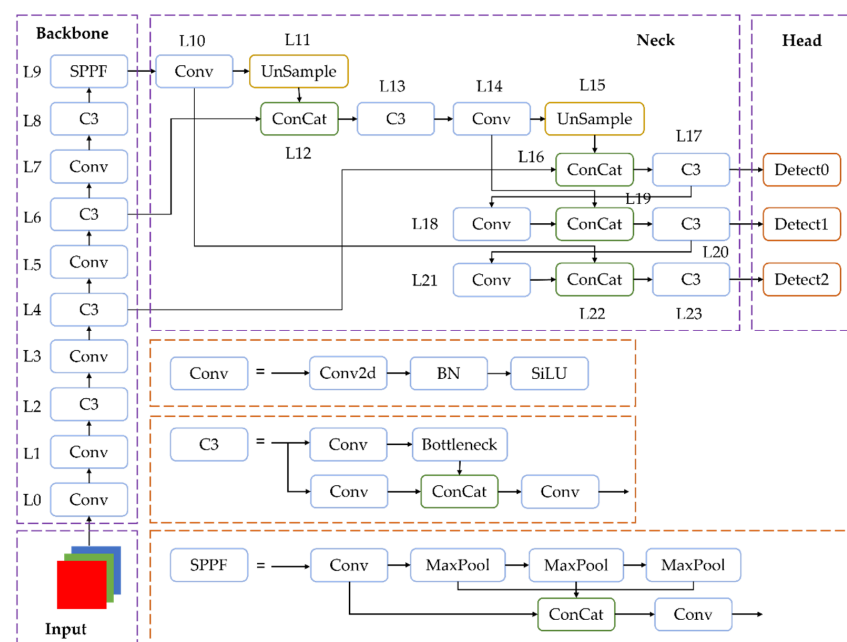
(d) Canopy fire from the UAV view

**Figure 1.** Training image samples in the dataset. (a,b) Surface fires and (c,d) canopy fires.

## 2.2. YOLOv5

As the network model continues to improve toward lightweight, the YOLO series incorporates effective methods from FCOS, RepVGG, and other networks to achieve better detection. YOLOv5 is currently a more mainstream target detection algorithm. YOLOv5s network is the network with the smallest depth and the smallest feature map width in the YOLOv5 series, and has faster recognition speed and higher accuracy.

In this paper, the forest fire classification detection model is improved based on version 6.0 of YOLOv5. The network structure is divided into four parts: Input, Backbone, Neck, and Prediction. The network structure of YOLOv5 is shown in Figure 2. First, in the Input part, the data are processed to increase the accuracy and discrimination of detection. Second, the Backbone part is mainly divided into Conv, C3, and SPPF module. Third, the Neck network adopts the feature pyramid structure of FPN (Feature Pyramid Network) and PANet (Pyramid Attention Network). FPN [30] transfers the semantic information from the deep layer to the shallow layer to enhance the semantic representation at different scales. PANet [28] builds on this and introduces a bottom-up path. This makes it easier to transfer the information from the bottom to the top and enhances the localization ability at different scales. Finally, the Prediction part is to use the loss function to calculate the position, classification, and confidence loss, respectively. YOLOv5 uses CIOU Loss to calculate the bounding box loss by default and considers the width-to-height ratio of the bounding box based on DIOU Loss [25]. It makes the target box regression more stable.

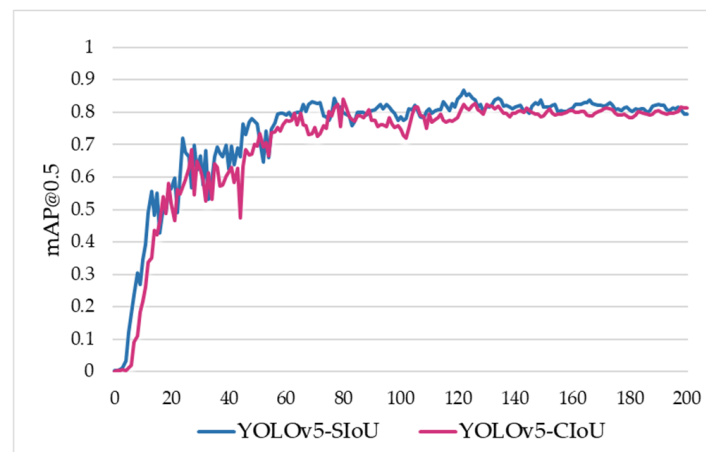


**Figure 2.** Structure of YOLOv5s model version 6.0.

### 2.3. SIoU Loss

The loss function plays an important role in model performance [31]. It is an important metric for measuring the difference between the predicted and true values of a deep convolutional neural network and is particularly important for a target detection algorithm. Therefore, the loss function determines how well the model parameters are trained.

The traditional target detection loss function such as CIoU Loss in YOLOv5 increases the consideration of aspect ratio based on the previous loss function, but does not consider the vector angle between the required regressions, i.e., the direction of mismatch between the required real frame and the prediction frame. In addition, this can lead to the problem of wandering of the prediction frame during the training process and reduce the experimental training speed. The definitions are as follows. The literature [26] proposes a new loss function SIoU Loss based on CIoU Loss, which does not only rely on the aggregation of the bounding box regression indicators and introduces directionality in the loss function cost to redefine the penalty indicators. As shown in Figure 3, the blue curve is the YOLOv5-SIoU training effect and the red curve is the YOLOv5-CIoU training effect. The experiments prove that SIoU converges faster and trains better compared to CIoU.



**Figure 3.** Comparison of the training effect of SIoU and CIoU based on YOLOv5 model.

SIoU loss function consists of four cost functions, including Angle cost, Distance cost, Shape cost, and *IoU* cost. The definitions are as follows.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

$$Loss_{CIoU} = 1 - IoU + \frac{\rho^2(A, B)}{c^2} + \alpha v \quad (2)$$

$$Loss_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (3)$$

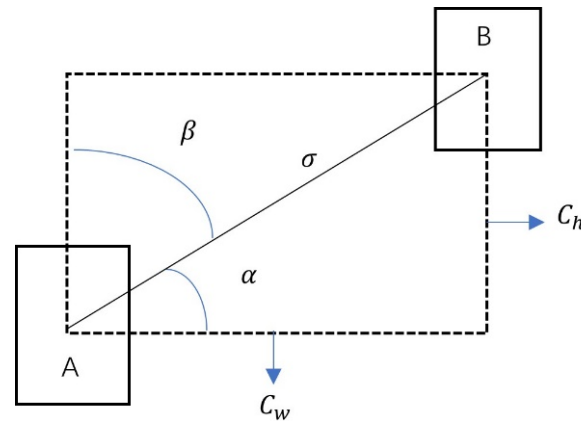
$$\Lambda = 1 - 2 \times \sin^2\left(\arcsin(x) - \frac{\pi}{4}\right) \quad (4)$$

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho_t}) \quad \gamma = 2 - \Lambda \quad (5)$$

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta \quad (6)$$

In the above equation, *IoU* is the *IoU* loss, where A represents the prediction frame and B represents the real frame.  $\rho$  is the distance between the centroids of box A and box B,  $c$  is the diagonal length of the minimum enclosing rectangle of box A and box B,  $v$  is the similarity of the aspect ratio of box A and box B, and  $\alpha$  is the influence factor of  $v$ .  $\Lambda$  is the angular loss that allows the prediction frame to move faster to the horizontal or vertical

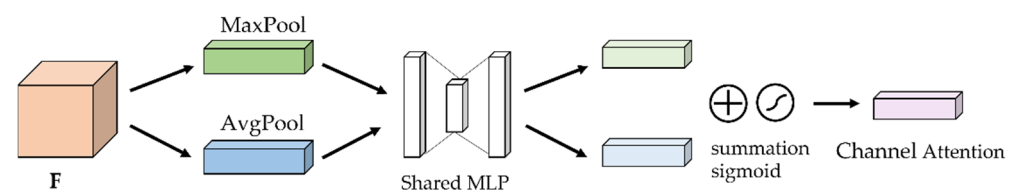
line where the true frame is located. The angle cost scheme in the loss function is shown in Figure 4.  $\Delta$  is the distance loss.  $\Omega$  is the shape loss.  $C_h$  is the height difference between the center point of the real frame and the predicted frame, and  $\sigma$  is the distance between the center point of the real frame and the predicted frame.



**Figure 4.** Angle cost scheme in the loss function.

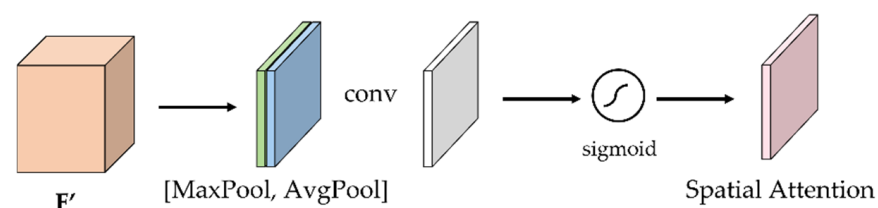
#### 2.4. CBAM

Forest fire targets generally have low pixels in the pictures. The target detection process is prone to missing information and missed false detection. Therefore, we add CBAM attention mechanism in the Backbone part of the network model. It can not only improve the detection accuracy for different types of forest fire targets, but also enhance the extraction of target features for detecting different types of forest fire targets, and reduce the cases of missed forest fires and mis-detected forest fire types. CBAM is used in feedforward convolutional neural networks. Combining the Channel Attention Module (CAM) and Spatial Attention Module (SAM), it is an uncomplicated and valid end-to-end general-purpose module. The structure of CBAM attention mechanism is divided into two parts. The first part is the channel attention module, as shown in Figure 5.



**Figure 5.** The structure of the Channel Attention Module.

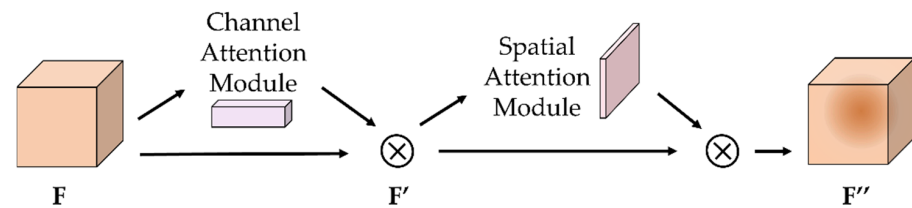
First, the given feature maps are used as inputs for global maximum pooling and mean pooling, respectively. The pooled two one-dimensional vectors are then fed into a shared Multi-Layer Perceptron (MLP). Then the two features' output from the MLP are subjected to summation operation and sigmoid function activation operation. The feature maps of the final generated channel attention module are used as the input features required for the spatial attention module, i.e., the second part, as shown in Figure 6.



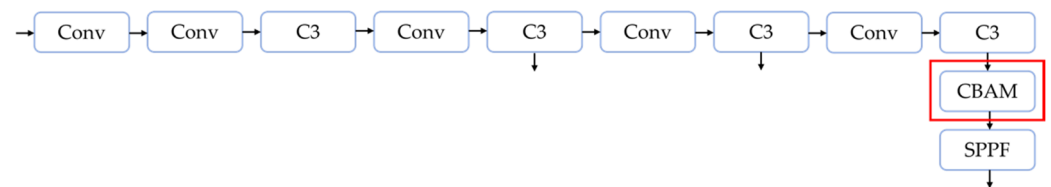
**Figure 6.** The structure of the Spatial Attention Module.

The global maximum pooling and global average pooling are performed on the feature maps output by the channel attention module. In addition, the two binary vectors after pooling are channel spliced. Then the convolution operation descending and sigmoid activation function operation is performed. The final feature map of the spatial attention module is obtained.

The CBAM attention mechanism module integrates the above two components. For the input feature maps, the CBAM module inferred the attention maps sequentially along both channel and spatial dimensions. The attention graph is then multiplied with the input feature graph to perform adaptive feature optimization. Its module structure is shown in Figure 7. Adding CBAM locations to the network is shown in Figure 8.



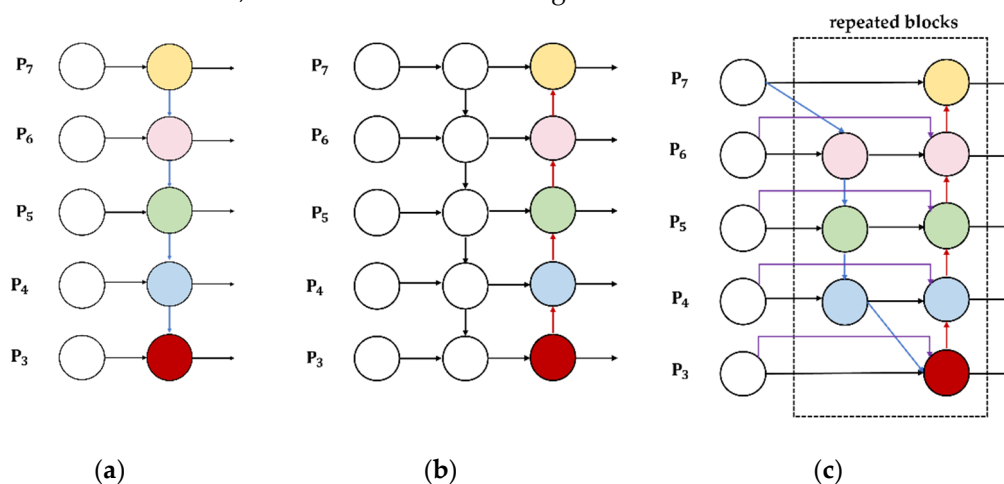
**Figure 7.** The structure of the CBAM module.



**Figure 8.** The location of the CBAM added in the network.

### 2.5. BiFPN

The difficulty of target detection lies in finding an efficient way to fuse multi-scale features [32]. YOLOv5 uses a Path Aggregation Network (PANet). An additional FPN bottom-up path aggregation network is added to the Feature Pyramid Network (FPN). We consider improving the PANet layer into a weighted bi-directional feature pyramid network (BiFPN). BiFPN constructs top-down and bottom-up bi-directional channels and repeats the same level to achieve simple and fast fusion and screening of forest fire features in different dimensions. BiFPN introduces weights in feature fusion compared to the previous equal treatment of different scale features. This can better balance the feature information of different scales of forest fire, prevents the loss of forest fire features, and improves the inspection of different types of forest fires. The structural design of FPN, PANet, and BiFPN is shown in Figure 9.

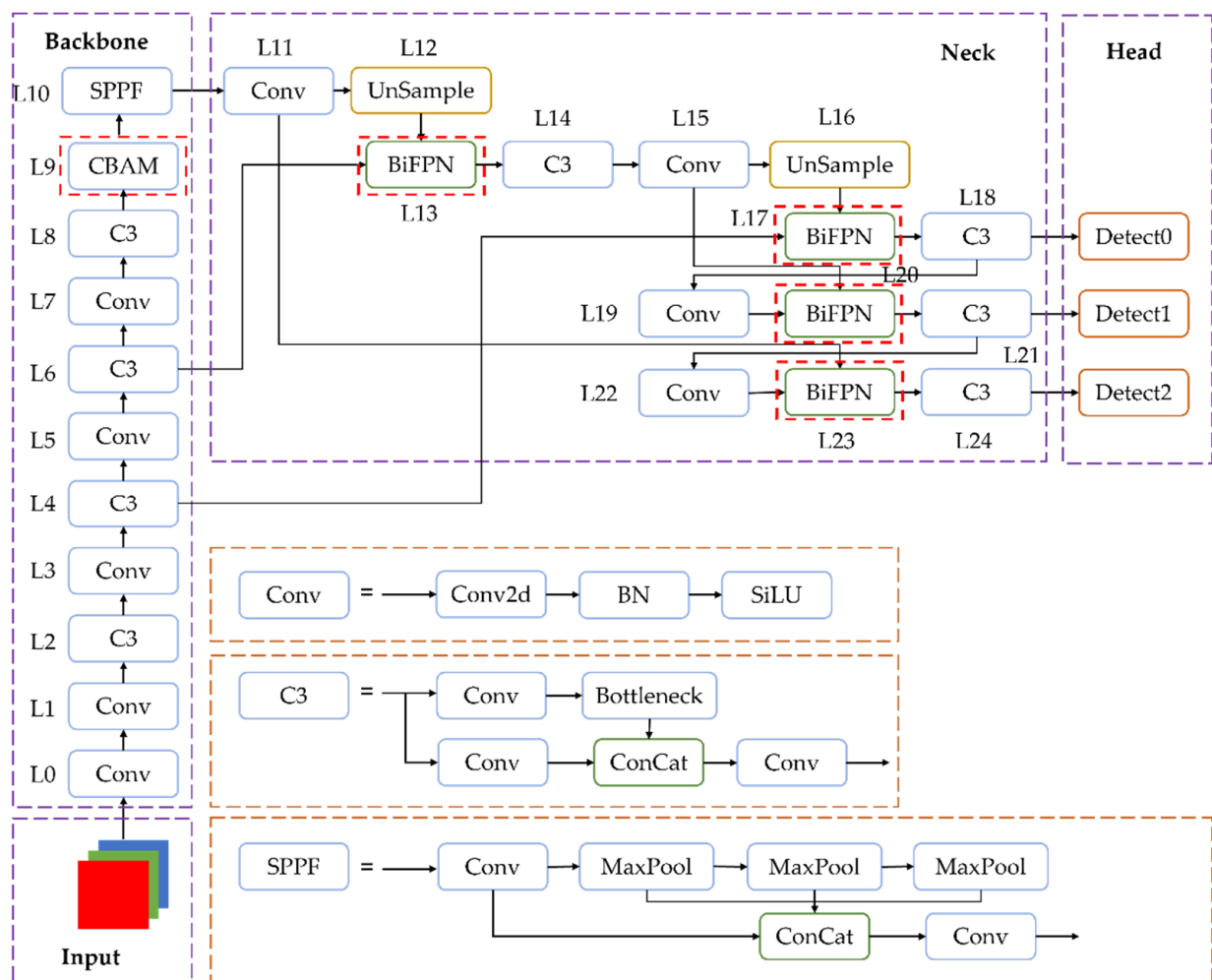


**Figure 9.** The structural design of FPN, PANet, and BiFPN. (a) FPN; (b) PANet; (c) BiFPN.

With the depth of the network layers, there is a certain degree of missing features when converting low-dimensional semantics to high-dimensional semantics. Therefore, it is necessary to fuse the features of different layers. In this paper, BiFPN is introduced in the Neck network layer for feature fusion to improve the model performance.

## 2.6. Improved Forest Fire Classification and Detection Model FCDM

Forest fire targets generally have the problem of too-low pixels in the pictures. Surface fires and canopy fires are difficult to distinguish after the fire has started for a period of time, and the target detection process is prone to missing information and false detection by omission. To address this problem, we improve the original YOLOv5s model. The improved forest fire classification and recognition model based on YOLOv5s version 6.0 in this paper is shown in Figure 10.



**Figure 10.** Improved Forest Fire Classification and Detection Model FCDM.

Firstly, in the input side part, the data input is subjected to adaptive image scaling, Mosaic data enhancement, and adaptive anchor frame calculation. The data are processed to increase the accuracy and discrimination of detection. Second, the Backbone part introduces CBAM attention module. Conv module encapsulates three functional modules of Conv2d, Batch Normalization (BN), and activation function (SiLU). The C3 module borrows the idea of CSPNet [33] module to improve the inference speed and maintain the recognition accuracy of model detection while reducing the computation. SPPF is proposed based on SPP [34], and the module uses multiple small-size pooled kernel cascade generation to fuse the feature maps of different sensory fields to further improve the operation speed.

CBAM improves the precision of target detection and enhances the extraction of detection target features. Third, the weighted Bi-directional Feature Pyramid Network (BiFPN) is introduced in the Neck network layer part to prevent feature loss and improve the inspection of different types of forest fires. It makes it easier to transfer the information from the bottom layer to the top and enhances the localization ability at different scales. Finally, the Prediction part uses the loss function to calculate the location, classification, and confidence losses, respectively. The bounding box loss function is changed to SIOU Loss. Directionality is added to the cost of the loss function, which enhances the training and inference of detection algorithms.

### 3. Results

#### 3.1. Training

The experiment conditions in this paper are shown in Table 1. Based on the default parameter values of YOLOv5 and continuous adjustment of experiments, the training parameters of the improved forest fire classification and detection model are set as shown in Table 2. The initial learning rate of the YOLOv5 model is 0.01, and the epoch is automatically adjusted with the effect of the mold validation set. The epoch of 200 is the best fit. The forest fire classification dataset is divided into training set, validation set, and test set in the ratio of 8:1:1. The details of the forest fire classification dataset are shown in Table 3.

**Table 1.** Experimental conditions.

Experimental Environment	Details
Programming language	Python 3.9
Operating system	Windows 11
Deep learning framework	Pytorch 1.12.1
GPU	NVIDIA GeForce RTX 3060

**Table 2.** Training parameters setting for improved forest fire detection model.

Training Parameters	Details
Epochs	200
batch-size	8
Img-size	640 × 640
Initial learning rate	0.01
Optimization algorithm	SGD

**Table 3.** Details of our dataset.

Dataset	Train	Val	Test
forest fire dataset	441	49	54
surface fire	242	27	31
canopy fire	197	22	25

#### 3.2. Model Evaluation

To verify the detection performance of the improved forest fire classification detection model in this paper, the Microsoft COCO evaluation metric [35] is used in this paper. This metric is widely used to evaluate target detection tasks and is currently recognized as one of the more authoritative standards in terms of target recognition.

The evaluation metrics use precision (P), recall (R), and mAP@0.5 to compare the performance of each model. Among them, mAP@0.5 is the average accuracy of each class of images when *IoU* is set to 0.5, and is the main evaluation metric for target detection

algorithms. The higher the mAP@0.5 value, the better the target detection model detects on a given dataset. The calculation equations are as follows:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$AP = \int_0^1 P(r)dr \quad (9)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (10)$$

$$Time = Pre - process + Inference + NMS \quad (11)$$

$$FPS = \frac{1}{Time} \quad (12)$$

In the above equation, *Precision* is the ratio of all correct predictions of the model in which the true value is a positive sample. *Recall* is the proportion of the model's correct predictions among the total positive samples. *P(r)* denotes the P–R curve, and the horizontal and vertical coordinates are recall and accuracy. *AP* is the average accuracy, which is represented in the P–R graph as the area enclosed by the horizontal and vertical coordinates. *Time* represents the time taken for the whole training and consists of three parts. *FPS* indicates how many forest fire images can be processed in one second.

Taking forest fire detection as an example, *TP* denotes the number of correctly predicted forest fire images. That is, if the detection target is forest fire, the model detection result is forest fire. *FP* means the detection target is non-forest fire and the model detection result is non-forest fire. *FN* means the detection target is forest fire and the model detection result is non-forest fire. Pre-process indicates the image pre-processing time. Inference indicates the training inference time. NMS, also known as non-maximum suppression, is the frame post-processing time.

The experimental procedure is described as follows. First, the original YOLOv5 detection model is trained using the forest fire classification dataset and evaluated using the test set. Then different loss functions are modified and different attention mechanisms are added for training. The best combination is selected based on the detection results. Attentional mechanisms are developing rapidly, and the more mainstream attentional modules are currently SE, CBAM, and GAM. In this experiment, these three typical attention modules were selected for experimental comparison. The SE (Squeeze-and-Excitation) attention module [36] solves the loss problem caused by the different importance occupied by different channels of the feature map in the convolutional pooling process, which can improve the accuracy of the forest fire detection network. The CBAM attention module considers more spatial attention on the basis of SE, and uses both tie pooling and maximum pooling to reduce the loss of forest fire information brought by pooling to a certain extent. The GAM (Global Attention Mechanism) [37] adopts a sequential channel-space attention mechanism and redesigns the CBAM sub-module to amplify the global dimensional interaction features while reducing the dispersion of forest fire information. These three attention mechanisms are the current mainstream modules. Finally, the Path Aggregation Network (PANet) layer is replaced with a weighted Bi-directional Feature Pyramid Network (BiFPN) to build a forest fire classification recognition model. The detection model was tested using the same test set. The mAP@0.5 results were obtained as shown in Table 4.

**Table 4.** Experimental results of the model.

Model	Forest Fire Detection (mAP@0.5)	Surface Fire Detection (mAP@0.5)	Canopy Fire Detection (mAP@0.5)	FPS	Time (ms)
YOLOv5s	0.830	0.791	0.868	58	17.2
YOLOv5s + Alpha-IoU	0.802	0.761	0.844	62	16.1
YOLOv5s + SIoU	0.851	0.805	0.896	63	15.9
YOLOv5s + SIoU + SE	0.851	0.791	0.901	63	15.9
YOLOv5s + SIoU + GAM	0.860	0.803	0.907	28	35.7
YOLOv5s + SIoU + CBAM	0.864	0.828	0.899	61	16.4
YOLOv5s + SIoU + GAM + BiFPN	0.861	0.793	0.908	56	17.9
YOLOv5s + SIoU + CBAM + BiFPN (FCDM, ours)	0.869	0.831	0.906	64	15.6

### 3.3. Detect Performance and Analysis

From the above results, we found that YOLOv5, as one of the current more advanced single-stage target detection models, has a better mAP@0.5 for forest fire classification recognition, but there is still room for improvement. The FPS value was 58 and the detection time was 17.2ms, the detection speed was low, and the detection was slow in real-time detection. In Experiments 2–3, the loss functions in YOLOv5s were changed to Alpha-IoU Loss and SIoU Loss, respectively. In Experiment 2, the mAP@0.5 values of forest fire inspection and forest fire classification detection were decreased to some extent. In addition, in Experiment 3, forest fire, surface fire, and canopy fire improved by 2.1%, 1.4%, and 2.8%, respectively. This indicates that SIoU Loss introduces directionality in the loss function cost, which improves the training and inference of forest fire detection to a larger extent.

Based on Experiment 3, three attention mechanisms were added for comparison in Experiments 4–6: SE, GAM, and CBAM, respectively. The GAM attention mechanism was better for canopy fire detection with concentrated distribution, improving by 1.1%, but there was a slight decrease for surface fire detection. In addition, the CBAM attention mechanism had some degree of improvement for all forest fire detection. Among them, it was better for the detection of surface fires with more scattered distribution and irregular shape, improving by 2.3%. The detection speed is faster and occurs in real-time detection; the FPS value is 61 and the detection duration is 16.4 ms.

Based on the addition of the attention mechanism, Experiments 7–8 improved the path aggregation network PANet layer into BiFPN. The experiments demonstrated that there was a slight improvement of map value after improving BiFPN. It indicated that BiFPN can fuse and filter different dimensional forest fire features to prevent feature loss and improve the detection of forest fire at different scales. The detection speed is faster and satisfies the requirements of forest fire in real-time detection with an FPS value of 64 and a detection time of 15.6 ms.

Finally, Experiment 8 with a more balanced enhancement effect was used as the final forest fire classification detection model selected in this paper, with mAP@0.5 values of 86.9%, 83.1%, and 90.6% for forest fire, surface fire, and canopy fire detection, respectively. Compared with the YOLOv5s model, the mAP@0.5 of this model improved by 3.9%, 4.0%, and 3.8%, respectively, indicating that this model is better for forest fire classification detection.

The effect of the YOLOv5 model and FCDM forest fire classification detection model is shown in Figures 11 and 12. Figure 11 shows the detection effect of forest fire training image samples with typicality. Figure 12 shows the detection effect of forest fire image samples from a UAV camera view. From the detection effect figure, the YOLOv5 model detection of the forest fire rectangular box position does not fit with the real box, and there will be the problem of missed detection, and the detection effect is poor. The FCDM model detection fits better with the real forest fire target frame, with better detection of forest fire types and less false detection and omission. It indicates that the forest fire classification detection model FCDM presented in this paper is better for forest fire classification detection.



(a) Surface fires' detection using YOLOv5



(b) Surface fires' detection using FCDM



(c) Canopy fires' detection using YOLOv5



(d) Canopy fires' detection using FCDM

**Figure 11.** The detection effect of YOLOv5 and FCDM model for typical forest fire images. (a,b) Surface fires and (c,d) canopy fires.



(a) Surface fires' detection using YOLOv5



(b) Surface fires' detection using FCDM



(c) Canopy fires' detection using YOLOv5



(d) Canopy fires' detection using FCDM

**Figure 12.** The detection effect of YOLOv5 and FCDM model for forest fire images from UAV viewpoint. (a,b) Surface fires and (c,d) canopy fires.

#### 4. Discussion

Forests are important for maintaining the ecological security of the planet. Uncontrolled forest fires pose a threat to forests and may have far-reaching effects. Over the past half-century, the area burned by forest fires in the western United States has increased tenfold each year [38]. And fires that are not disposed of in a timely manner can pose a significant safety hazard. As the technical methods in the field of target detection continue to mature, it is of some practical significance to apply them to detect and identify fire types and take corresponding suppression measures in a timely manner. Forest fire detection techniques are constantly being improved, but it is difficult to distinguish forest fire types based on pictures. The detection of different types of forest fires has not been well-studied.

Accordingly, for the above reasons, forest fire classification detection technology needs further research and development. Through experiments, we found that the YOLOv5 model has better detection results in identifying forest fires. However, surface fires and canopy fires are difficult to distinguish after a period of time after the fire starts, and the pictures have the problem of low pixels. The target detection process is prone to missing information and false detection by omission. Therefore, we considered changing the bounding box loss function to SIOU Loss to improve the training and inference of the detection algorithm. We also introduced the CBAM attention mechanism into the network to improve the classification recognition accuracy. We also improved the PANet layer as BiFPN to enhance the inspection of different types of forest fires.

However, the forest fire classification and detection model proposed in this paper still has some shortcomings. Therefore, we consider further optimization of this model. First, different detection models and attention mechanisms will produce different detection effects for different forest fire types with different distribution and target shapes. For example, the CBAM attention mechanism has a certain improvement for all types, while the GAM only has a higher improvement for canopy fire detection. Further search for models with better detection for individual forest fire types is considered for integration. Secondly, forest fire types are difficult to distinguish after a period of time from the beginning of the fire, and surface fires tend to turn into canopy fires as they increase in intensity. Further photographing or searching for images of the early stage of fire is considered to extend the dataset and enhance the reliability of the dataset.

The experiment results show that the forest fire classification and recognition model proposed in this paper has good application prospects. In practical applications, cameras are considered to be loaded on UAVs or helicopters. Compared with forest fire identification techniques using a large number of sensors, different fire-fighting measures can be taken in a timely manner according to the detection results while saving costs. Therefore, we believe that the forest fire classification and detection model proposed in this paper is more advantageous.

#### 5. Conclusions

The frequent occurrence of global forest fires, the increased difficulty of fire control, and the lack of timely disposal have caused great property losses and serious safety problems. Therefore, it has a certain development space and practical significance to detect and identify fire types and take corresponding fire extinguishing measures in a timely manner.

In this paper, an improved forest fire classification and detection model FCDM based on YOLOv5 is presented. Firstly, the bounding box loss function is changed to SIOU Loss, and the directionality is introduced in the loss function cost to improve the training and inference of the detection algorithm. Secondly, the CBAM attention mechanism module is introduced into the network to enhance the classification identification accuracy. Then, the PANet layer is improved as a weighted BiFPN to enhance the inspection of different types of forest fires. The experimental results show that the forest fire classification and detection model proposed in this paper outperforms the YOLOv5 algorithm in terms of forest fire, surface fire, and canopy fire detection performance.

In further subsequent studies, we will keep testing the model in this paper to further improve its detection performance and consider its value in practical applications. Additionally, we will think about combining several single detection models that are effective for particular types of forest fires to see if their fusion models increase detection for all species in comparison to a single model. To better deploy the model in practical applications, we will further develop the detection performance of the model and study the light weighting of the model. We hope to combine more advantages in forest fire detection techniques in future research, so that forest fire classification and detection techniques can continue to develop and innovate.

**Author Contributions:** Q.X. devised the programs and drafted the initial manuscript. H.L. and F.W. designed the project and revised 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), the Jiangsu Modern Agricultural Machinery Equipment and Technology Demonstration and Promotion Project (NJ2021-19), the National Natural Science Foundation of China (32101535), the Jiangsu Postdoctoral Research Foundation (2021K112B), and the Nanjing Modern Agricultural Machinery Equipment and Technological Innovation Demonstration Projects (No. NJ [2022]09).

**Data Availability Statement:** Available.

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

## References

1. Gong, L. Effect and Function of Forest Fire on Forest. *For. Sci. Technol. Inf.* **2011**, *3*, 46–48.
2. Tang, S.; Li, H.; Shan, Y.; Xiao, Y.; Yi, S. Characteristics and Control of Underground Forest Fire. *World For. Res.* **2019**, *32*, 42–48.
3. Zhu, M.; Feng, Z.K.; Lin, H.U. Improvement of forest surface fire spread models. *J. Beijing For. Univ.* **2005**, *S2*, 145–148.
4. Hao, Q.; Huang, X.; Luo, C.; Yang, Y. Realization and Application of Mountain Fire Infrared Radar Alarming System of Transmission Line. *Guangxi Electr. Power* **2015**, *38*, 67–71.
5. Yuan, C.; Liu, Z.; Zhang, Y. Vision-based Forest Fire Detection in Aerial Images for Firefighting Using UAVs. In Proceedings of the International Conference on Unmanned Aircraft Systems, Arlington, VA, USA, 7–10 June 2016.
6. Prosekov, A.Y.; Rada, A.O.; Kuznetsov, A.D.; Timofeev, A.E.; Osintseva, M.A. Environmental monitoring of endogenous fires based on thermal imaging and 3D mapping from an unmanned aerial vehicle. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2022; Volume 981, p. 042016.
7. Muid, A.; Kane, H.; Sarasawita IK, A.; Evita, M.; Aminah, N.S.; Budiman, M.; Djamil, M. *Potential of UAV Application for Forest Fire Detection*; IOP Publishing Ltd.: Bristol, UK, 2022.
8. Hua, L.; Shao, G. The progress of operational forest fire monitoring with infrared remote sensing. *J. For. Res.* **2017**, *28*, 215–229. [[CrossRef](#)]
9. Long, X.M.; Zhao, C.M.; Ding, K.; Yang, K.Y.; Zhou, R.L. A Method of Improving Geometric Accuracy of Forest Fire Monitoring by Satellite Imagery Based on Terrain Features. *Remote Sens. Inf.* **2016**, *31*, 89–94.
10. Ganesan, P.; Sathish, B.S.; Sajiv, G. A comparative approach of identification and segmentation of forest fire region in high resolution satellite images. In Proceedings of the 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave), Coimbatore, India, 29 February–1 March 2016.
11. Lin, J.; Lin, H.; Wang, F. STPM\_SAH: A Small-Target Forest Fire Detection Model Based on Swin Transformer and Slicing Aided Hyper Inference. *Forests* **2022**, *13*, 1603. [[CrossRef](#)]
12. Xue, Z.; Lin, H.; Wang, F. A Small Target Forest Fire Detection Model Based on YOLOv5 Improvement. *Forests* **2022**, *13*, 1332. [[CrossRef](#)]
13. Qian, J.; Lin, H. A Forest Fire Identification System Based on Weighted Fusion Algorithm. *Forests* **2022**, *13*, 1301. [[CrossRef](#)]
14. Yuan, C.; Liu, Z.; Zhang, Y. UAV-based forest fire detection and tracking using image processing techniques. In Proceedings of the 2015 International Conference on Unmanned Aircraft Systems (ICUAS), Denver, CO, USA, 9–12 June 2015; pp. 639–643.
15. Han, X.F.; Jin, J.S.; Wang, M.J.; Jiang, W.; Gao, L.; Xiao, L.P. Video fire detection based on Gaussian Mixture Model and multi-color features. *Signal Image Video Process.* **2017**, *11*, 1419–1425. [[CrossRef](#)]
16. 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](#)]
17. Mahmoud, M.A.I.; Ren, H. Forest fire detection and identification using image processing and SVM. *J. Inf. Process. Syst.* **2019**, *15*, 159–168.
18. Lin, H.; Tang, C. Analysis and Optimization of Urban Public Transport Lines Based on Multiobjective Adaptive Particle Swarm Optimization. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 16786–16798. [[CrossRef](#)]

19. Chen, Y.; Zhang, Y.; Xin, J.; Wang, G.; Mu, L.; Yi, Y.; Liu, H.; Liu, D. UAV image-based forest fire detection approach using convolutional neural network. In Proceedings of the 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 19–21 June 2019; pp. 2118–2123.
20. Barmpoutis, P.; Dimitropoulos, K.; Kaza, K.; Grammalidis, N. Fire detection from images using faster R-CNN and multidimensional texture analysis. In Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-2019), Brighton, UK, 12–17 May 2019; pp. 8301–8305.
21. Li, P.; Zhao, W. Image fire detection algorithms based on convolutional neural networks. *Case Stud. Therm. Eng.* **2020**, *19*, 100625. [\[CrossRef\]](#)
22. Majid, S.; Alenezi, F.; Masood, S.; Ahmad, M.; Gündüz, E.S.; Polat, K. Attention based CNN model for fire detection and localization in real-world images. *Expert Syst. Appl.* **2022**, *189*, 116114. [\[CrossRef\]](#)
23. Barisic, A.; Car, M.; Bogdan, S. Vision-based system for a real-time detection and following of UAV. In Proceedings of the 2019 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED UAS), Cranfield, UK, 25–27 November 2019; pp. 156–159.
24. Ultralytics-Yolov5. Available online: <https://github.com/ultralytics/yolov5> (accessed on 5 June 2022).
25. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12993–13000.
26. Gevorgyan, Z. SIoU Loss: More Powerful Learning for Bounding Box Regression. *arXiv* **2022**, arXiv:2205.12740.
27. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
28. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
29. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2020; pp. 10781–10790.
30. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
31. Lin, H.; Tang, C. Intelligent Bus Operation Optimization by Integrating Cases and Data Driven Based on Business Chain and Enhanced Quantum Genetic Algorithm. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 9869–9882. [\[CrossRef\]](#)
32. Lin, H.; Han, Y.; Cai, W.; Jin, B. Traffic Signal Optimization Based on Fuzzy Control and Differential Evolution Algorithm. *IEEE Trans. Intell. Transp. Syst.* **2022**, 1–12. [\[CrossRef\]](#)
33. Wang, C.Y.; Liao, H.Y.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 390–391.
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [\[CrossRef\]](#)
35. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
36. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
37. Liu, Y.; Shao, Z.; Hoffmann, N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. *arXiv* **2021**, arXiv:2112.05561.
38. Abatzoglou, J.T.; Battisti, D.S.; Williams, A.P. Projected increases in western US forest fire despite growing fuel constraints. *Commun. Earth Environ.* **2021**, *2*, 1–8. [\[CrossRef\]](#)