

Article

# An Improved Forest Fire and Smoke Detection Model Based on YOLOv5

Junhui Li <sup>1</sup>, Renjie Xu <sup>2</sup> and Yunfei Liu <sup>1,\*</sup> 

<sup>1</sup> The College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China; lijunhui@njfu.edu.cn

<sup>2</sup> Department of Computing and Software, McMaster University, Hamilton, ON L8S 4L8, Canada

\* Correspondence: lyf@njfu.edu.cn; Tel.: +86-25-8542-7827

**Abstract:** Forest fires are destructive and rapidly spreading, causing great harm to forest ecosystems and humans. Deep learning techniques can adaptively learn and extract features of forest fires and smoke. However, the complex backgrounds and different forest fire and smoke features in captured forest fire images make detection difficult. Facing the complex background of forest fire smoke, it is difficult for traditional machine learning methods to design a general feature extraction module for feature extraction. Deep learning methods are effective in many fields, so this paper improves on the You Only Look Once v5 (YOLOv5s) model, and the improved model has better detection performance for forest fires and smoke. First, a coordinate attention (CA) model is integrated into the YOLOv5 model to highlight fire smoke targets and improve the identifiability of different smoke features. Second, we replaced YOLOv5s original spatial pyramidal ensemble fast (SPPF) module with a receptive field block (RFB) module to enable better focus on the global information of different fires. Third, the path aggregation network (PANet) of the neck structure in the YOLOv5s model is improved to a bi-directional feature pyramid network (Bi-FPN). Compared with the YOLOv5 model, our improved forest fire and smoke detection model at mAP@0.5 improves by 5.1%.

**Keywords:** forest fire smoke; forest fire detection; YOLOv5; coordinate attention; receptive field block module; path aggregation network; bi-directional feature pyramid network



**Citation:** Li, J.; Xu, R.; Liu, Y. An Improved Forest Fire and Smoke Detection Model Based on YOLOv5. *Forests* **2023**, *14*, 833. <https://doi.org/10.3390/f14040833>

Academic Editor: Paul-Antoine Santoni

Received: 23 March 2023

Revised: 15 April 2023

Accepted: 16 April 2023

Published: 18 April 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

At present, international forest fire prevention technologies mainly include forest protection aircraft, infrared remote sensing, and satellite monitoring systems. Many studies used remote sensing satellite data for near-real-time monitoring and fire detection. For example, Ebrahim Ghaderpour [1], Eunna Jang [2], and Amy Marsha [3] used satellite remote sensing data to monitor fires and other disturbances in near real-time. Although some of these technologies are reliable, they require the use of high-altitude satellites and are complicated to construct; some technical solutions require high investment in infrastructure and maintenance costs. It is difficult to meet the actual needs of forest resources monitoring in China. Currently, the monitoring methods used in China mainly include ground patrol, lookout monitoring, aerial patrol, etc. The first two methods are inefficient, with low automation, and difficult to realize real-time monitoring; the latter method is costly and not easy to realize all-weather monitoring. With the progress in computer science and machine vision [4], the trend of replacing the traditional fire warning system with a computer application system has gradually become obvious. According to different objects, fire detection methods can be divided into two categories [5]: flame detection and smoke detection. Since the smoke generated by forest fires is visible before the fire, video smoke detection is receiving more and more attention in forest fire prevention. The traditional video smoke detection method mainly combines static features and dynamic features for smoke identification. Typical features of smoke include color, texture, motion direction, etc. Genoves [6] et al. studied the color characteristics of smoke in YUV space.

Yuan [7] proposed a method to extract smoke texture features using LBP and LBPV variance. Xue [8] used spatial wavelet transform to monitor the translucency of smoke. Qian [9] used optical flow calculation to calculate the motion features of smoke. The method of detecting smoke in video sequences based on saliency is proposed by Xue [10]. Different methods can achieve good performance in a specific image dataset. However, due to the poor robustness of the algorithms, the performance tends to be poor in different image datasets, and these methods are difficult to eliminate the complex interference in real practical and engineering applications. Most conventional smoke detection algorithms [11] involve a pattern recognition process, which includes feature extraction and classification, manual extraction of features, and the design of recognizers. After extracting candidate regions, smoke recognition is performed using static and dynamic smoke features. Although some progress has been made compared to detectors, most algorithms are only effective for the rapid detection of fires, making it difficult to extract the most essential features of smoke, and the detection process is relatively inefficient. Lin Ji et al. [12] used HSV (hue, saturation, value) color space and calculated the local area signal-to-noise ratio of consecutive frames to extract the dynamic change characteristics of smoke to detect the smoke occurrence area, and then combined with LBP (local binary patterns) texture features to distinguish the smoke area using support vector machine. This model has a high average detection accuracy on its test dataset, but its false alarm rate is high. Ji Lin et al. [13] proposed TDFF (triple multi-feature local binary patterns and derivative gabor feature fusion) smoke detection algorithm, which captures clearer texture features and optimized image edge grayscale information, and this algorithm has a higher average detection rate on the dataset in the paper. The average detection rate of this algorithm on the dataset in the paper is higher, and the false alarm rate is higher.

For forest fire detection, some scholars use color space rules for flame identification, which has a good recognition effect; however, extracting flame features by color space rules only is more one-sided, and some features are lost, which makes the results limited. In particular, Gong Chen et al. [14] designed a flame detection algorithm based on RGB and HSL filters; Bakri et al. [15] formulated rules in RGB and YCbCr color space, respectively, and separated the fire pixels satisfying the rules from the background; Wang et al. [16] proposed a fire recognition model based on the blue channel, and determined the key threshold by plotting the ROC curve of a large number of sample images Ganesan et al. [17] proposed an improved fuzzy C-mean clustering method, and conducted experiments on RGB and CIELab color spaces, which has a better performance in forest fire detection and segmentation of high-resolution satellite images, but the method is limited by the image quality.

Many other scholars use multi-feature flame detection methods with more comprehensive flame feature extraction, but they are suitable for indoor fire detection and have not yet been tested in open and complex environments such as forests. Wang et al. [18] proposed a fire detection model based on the fire color component, using the similarity of continuous video frames, and proposed a tracking algorithm based on the flame motion characteristics of the detection region, which achieved satisfactory results; Shi et al. [19] used the RGB-HIS model to detect the region similar to the flame, and detected the flame by identifying the change in the center of mass of continuous images, so as to distinguish the flame-like interference sources (such as towels, flashlight light, reflective metal, strong natural light, etc.); Gunawaardena et al. [20] used adaptive background subtraction to detect foreground moving objects, determined whether the detected foreground object was a flame by color rules, modeled and classified fire pixels using YCbCr color space, analyzed fire candidate regions in the time domain to detect flame flicker and distinguished flames from interferers.

Forest fire prevention requires “hitting early, small and hard”, so smoke and forest fire detection require not only real-time algorithm but also high detection accuracy. The existing smoke and forest fire detection algorithms do not fully meet the requirements of smoke detection, and the main challenges are as follows.

(1) The background of forest fires to be detected is very complex. There are not only trees and other obscurant, but also a class of smoke interference, light changes, etc. (2) The characteristics of smoke are constantly changing. Smoke is a non-rigid object whose shape, color, and texture can easily change over time, making it difficult to extract the most basic features of smoke. (3) High requirements are needed for real-time detection. In order to ensure detection accuracy, it is difficult to avoid the problem of increasing the number of network layers and increasing the amount of computation. (4) It is difficult to obtain real forest fire samples. Most of the existing samples are artificially simulated fire smoke, which has the problems of single background and imbalance.

Although the traditional forest fire and smoke detection methods have achieved certain results, the following problems still exist (1) poor accuracy. Feature extraction needs to rely on professional knowledge for manual feature selection, and there is unreasonable feature design and interference from external factors such as lighting and background, resulting in poor recognition results. (2) Poor robustness. Traditional forest fire and smoke detection algorithms usually perform well only in a certain scene and poorly in other scenes.

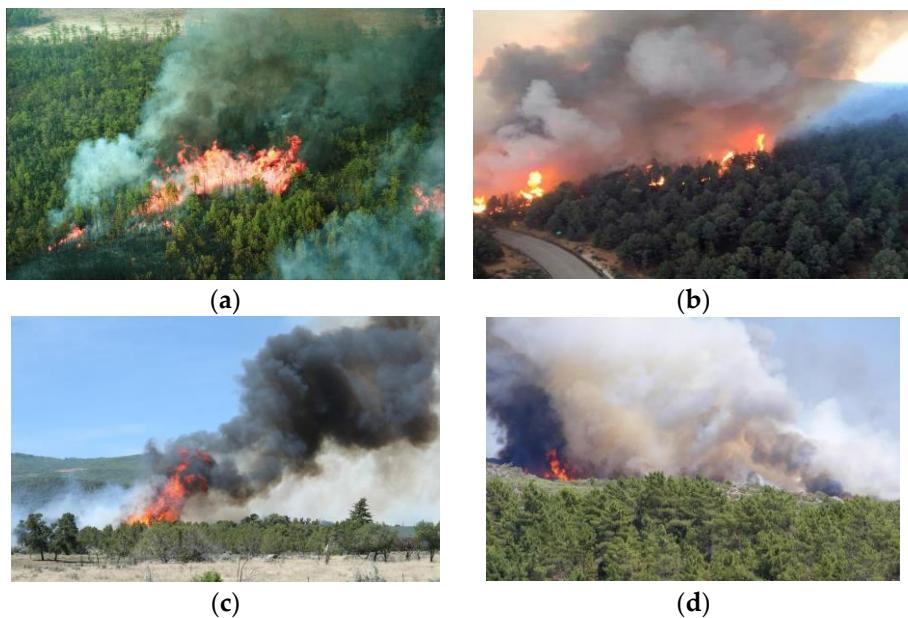
In this paper, an improved forest fire and smoke detection model based on YOLOv5 [21] is proposed. First, the CA attention mechanism is added to the backbone part of the YOLOv5 model to help the model better locate and identify forest fire and smoke targets in different contexts and improve the model's detection capability for various types of fires and smoke. Second, the SPPF module in the YOLOv5 model is replaced with RFB [22] to enable better focus on the global information of various forest fire smoke in the picture. Finally, the improved YOLOv5 model can better balance the information at different scales by improving the PANet [23] structure of the neck layer species to a Bi-FPN [24] structure.

The rest of the paper is organized as follows. In Section 2, we not only describe the forest fire and smoke dataset and model evaluation indicators used in our experiments but also detail the structure of our forest fire and smoke detection model. In Section 3, we show the configuration used for the experiments and the settings of some of the main training parameters. In addition, the gain of the CA attention module, RFB module, and Bi-FPN on forest fire and smoke detection is demonstrated by ablation experiments. In Section 4, our forest fire smoke detection model is discussed and analyzed. Section 5 summarizes the whole work and provides an outlook for future work.

## 2. Materials and Methods

### 2.1. Datasets

The learning effectiveness of a deep learner on target features is highly dependent on the annotation of the dataset. Therefore, the quality of the dataset has a very strong relationship with the effectiveness of model performance. First, we wrote a crawler program in python to collect forest fire and smoke images from the Internet. In addition, the high-quality forest fire and smoke images were manually screened. We removed the images of fire without smoke to make the model more robust. Images that are not forest fires and smoke are also filtered out. Secondly, we annotated the dataset by using tags to ensure that our model can identify forest fires and smoke. Finally, we produced a total of 450 images from the forest fire and smoke dataset. Representative images of each type in the dataset are shown in Figure 1.

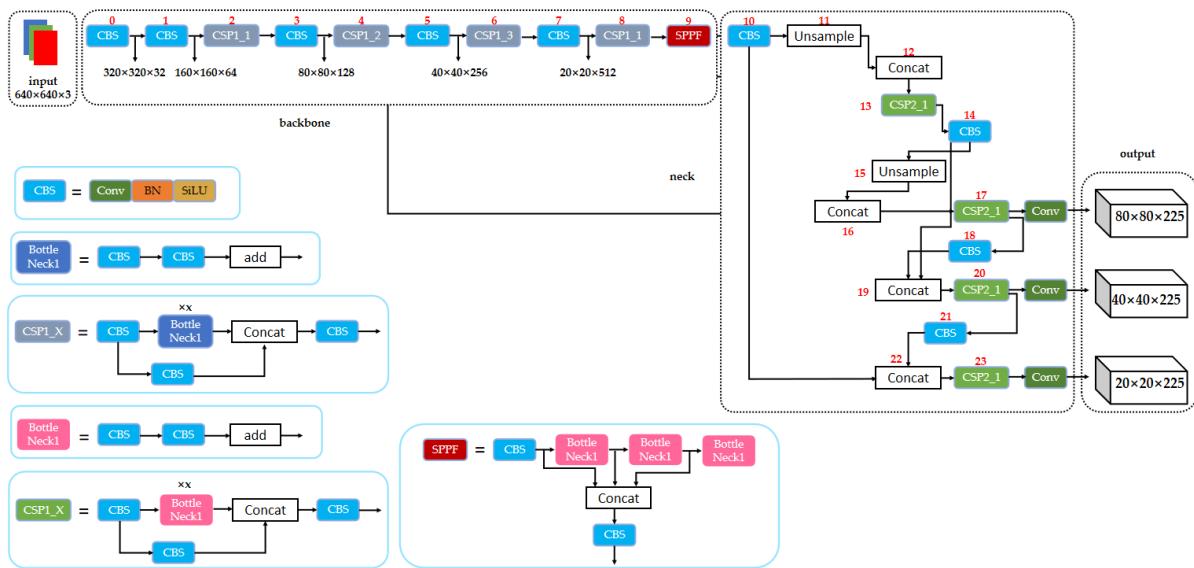


**Figure 1.** Forest fire smoke in a complex. (a,b) indicates that the fire is spreading, and smoke is present; (c,d) indicates that the fire is small, and smoke is present. context.

## 2.2. YOLOv5

YOLO is a one-stage detection model that balances speed and accuracy and is one of the best target detection models. YOLOv5 is an advanced target detection network model introduced by the Ultralytics LLC team in 2020. The image inference speed of the YOLOv5s model reaches 455 Frames Per Second (FPS), which is widely used by a large number of scholars with this advantage.

The YOLOv5 model has four parts: input, backbone, neck, and prediction. In this paper, we propose a forest fire smoke detection model that is based on the improved YOLOv5s model version 6.1. The structure of the YOLOv5s model in version 6.1 is shown in Figure 2.



**Figure 2.** The structure of YOLOv5s. The numbers represent the modules needed to travel from the input image to the final output image.

The input section of YOLOv5 consists of three main parts: image resizing, mosaic data augmentation, and adaptive anchor box calculation. Image resizing allows original

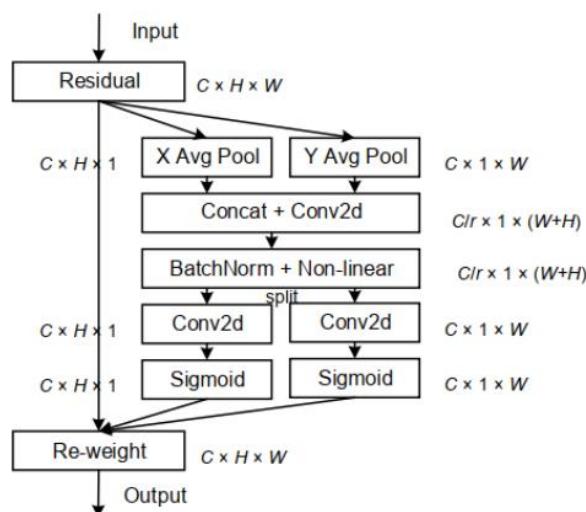
images of different lengths and widths to be scaled uniformly to a standard size by adding a minimum of black borders (so-called padding). Mosaic data augmentation in YOLOv5 defaults to Mosaic4 data augmentation, which allows four images to be randomly selected and combined to enrich the background of an image. The adaptive anchor frame calculation builds on the initial anchor frame by comparing the output prediction frame with the real frame, and the difference is calculated and then updated in reverse, iterating the parameters continuously to obtain the most suitable anchor frame.

In YOLOv5's backbone and neck, CSP1 and CSP2 are designed based on CSPNet [25] and have two different Bottleneck Cross Stage Partial (CSP) structures. The CSP module not only improves the learning performance of the neural network for the target features but also reduces the amount of parameter computation. CSP1 is used in the Backbone for feature extraction, and CSP2 is used in the Neck for feature fusion. In the Backbone, apart from the CSP1 module, there is also a Spatial Pyramid Pooling-Fast (SPPF) module, which is used to extract global information about the target. The SPPF module is equivalent to Spatial Pyramid Pooling (SPP), but it runs faster than the SPP module. In addition, the PANet structure is applied in the Neck layer, which not only fuses the extracted semantic features with the positional features but also fuses the features of the backbone layer with the detection layer, allowing the model to acquire richer feature information. Finally, the prediction network consists of three detection layers, with feature mean average precision (mAP)@0.5s of different sizes used to detect target objects of different sizes.

### 2.3. Coordinate Attention

Position information is essential for capturing the structure of objects in visual tasks. In traditional lightweight networks, most attention uses squeeze and excite (SE) attention [26]. However, they only consider inter-channel information and ignore location information. In order to take into account both inter-channel information as well as location information, this paper chooses to Coordinate Attention, which is a relatively new attention proposed by Qibin Hou et al. [27] in 2021.

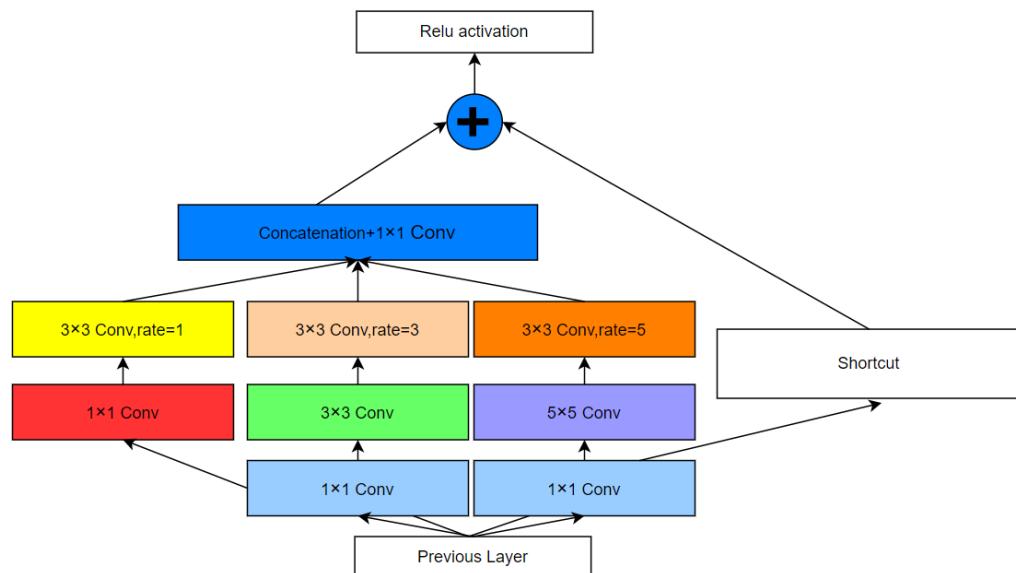
Coordination concerns enable mobile networks to focus major overhead on large areas and reduce computational complexity by embedding location information in channel information. In order to mitigate the loss of positional information due to the 2D global pooling, coordinate attention decomposes channel attention into two parallel 1D feature encoding processes to integrate spatial coordinate information effectively into the generated attention. Coordinate attention encodes channel relations and remote dependencies through two steps: coordinate information embedding and coordinate attention generation. The specific structure of Coordinate Attention is shown in Figure 3.



**Figure 3.** The structure of Coordinate Attention. C represents the number of channels; H represents the height; W represents the width.

#### 2.4. Receptive Field Block Net

Receptive Field Block (RFB) was presented by Liu S et al. [28] in 2018. The RFB was inspired by simulating the perceptual wilderness of human vision to enhance the feature extraction capabilities of the network. Construction of the RFB module by combining multiple branches with different kernels and dilated convolution layer. Multiple kernels are analogous to the pRFs of varying sizes. At the same time, dilated convolution layer assigns each branch with individual eccentricity to simulate the ratio between the size and eccentricity of the pRF. With a concatenation and  $1 \times 1$  convolution in all the branches, the final spatial array of RF is produced, which is similar to that in human visual systems. The structure of the RFB borrows ideas from Inception by adding a cavity convolution [29] to Inception, which effectively increases the receptive field. The structure of the RFB is shown in Figure 4.



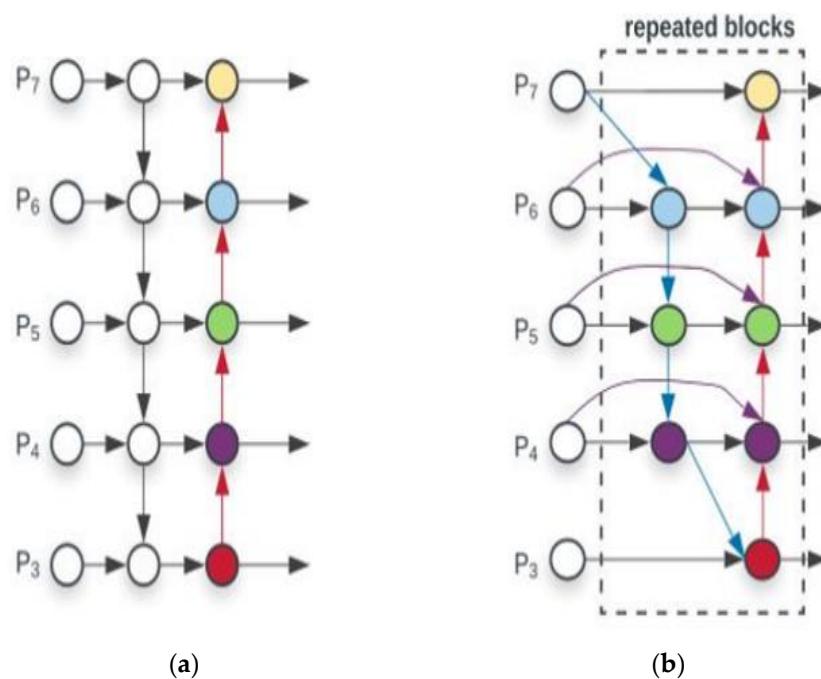
**Figure 4.** The structure of the RFB.

The RFB structure has two main features: firstly, a multi-branched structure consisting of convolutional layers with convolutional kernels of different sizes. Secondly, the introduction of dilated convolutional layers, whose main role is to increase the receptive field.

#### 2.5. Bi-Directional Feature Pyramid Network

Bidirectional Feature Pyramid Network (Bi-FPN) is a modified PANet structure proposed by Tan M et al. [30] in 2020. The structure of the PANet and BiFPN is shown in Figure 5.

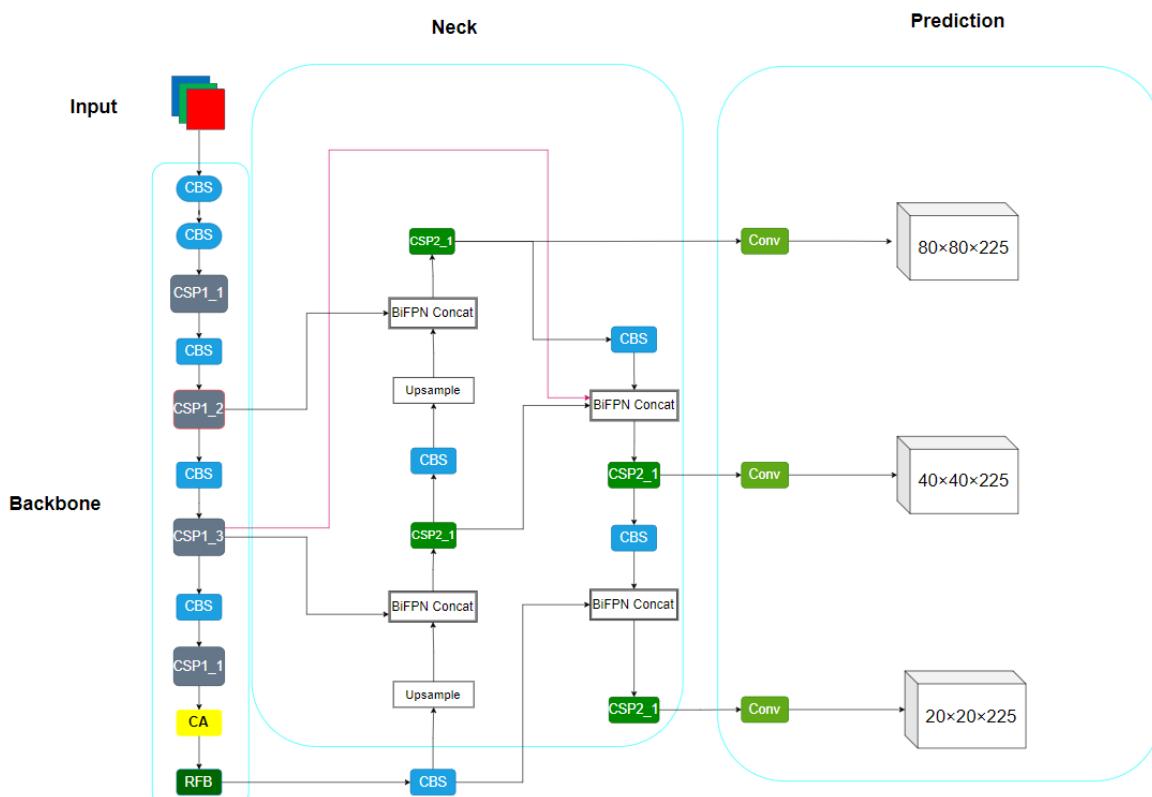
Based on the FPN, the PANet adds a fusion path from the bottom to the top, which enhances the multi-scale fusion information of the FPN [31]. Although the PANet structure used in YOLOv5 is more efficient, it will result in higher computational costs. Bi-FPN improves the PANet: first, remove intermediate nodes in the PANet. Secondly, a jump connection is added between input and output nodes of the same size. Since they are at the same layer, more features are fused without adding much computational cost. Bi-FPN has fewer parameters than PANet, allowing for a higher level of diagnosis-specific fusion.



**Figure 5.** (a) for PANet, (b) for BiFPN. The color represents the multi-scale feature representation of the previous P3-P7 fusion.

## 2.6. Our Improved Model

The structure of our proposed improved forest fire smoke detection model is shown in Figure 6. Firstly, we added coordinate attention behind layer 8 of the YOLOv5s model. Then, we replaced the original SPPF module of YOLOv5 with an RFB module. Finally, we modified the PANet structure of YOLOv5 to a Bi-FPN structure.



**Figure 6.** The structure of improved forest fire smoke detection model.

## 2.7. Model Evaluation

In this paper, PASCAL VOC, one of the commonly used evaluation criteria in target detection tasks [32–34], is used to evaluate the model. The evaluation metric of PASCAL VOC is measured by mAP@0.5. The mAP@0.5 is the mean of the average accuracy when the Intersection over Union (IOU) is set to 0.5. In order to calculate mAP@0.5, we need to use precision and recall. The calculation formula is as follows.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (1)$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (2)$$

$$\text{AP} = \int_0^1 \text{P}(r)dr \quad (3)$$

$$\text{mAP} = \sum_{i=1}^C \text{AP}_i/C \quad (4)$$

In those calculation formulas, TP = True Positive, FP = False Positive, and FN = False Negative.  $P(r)$  is a trial of the function corresponding to the curve of precision and recall, with precision as the  $x$ -axis and recall as the  $y$ -axis. AP is an area on the curve enclosed by Precision and Recall. The mAP@0.5 is the average of the AP values for all categories.

## 3. Results

### 3.1. Training

The model testing environment in this paper is shown in Table 1. The main parameter settings for the training of forest fire and smoke detection models in this paper are shown in Table 2. We divide the forest fire and smoke dataset of 450 images into a training set, validation set, and test set by 8:1:1. The training and validation sets are used for model training, and the test set is used for model testing. The details of the forest fire and smoke datasets are shown in Table 3.

**Table 1.** Model test environment.

| Test Environment        | Details                 |
|-------------------------|-------------------------|
| Programming language    | Python 3.8              |
| Operating system        | Windows 11              |
| Deep learning framework | Pytorch 1.8.2           |
| GPU                     | NVIDIA GeForce GTX 3060 |

**Table 2.** Training parameters for forest fire and smoke detection models.

| Training Parameters    | Details   |
|------------------------|-----------|
| Epochs                 | 200       |
| batch-size             | 16        |
| img-size (pixels)      | 640 × 640 |
| Optimization algorithm | SGD       |
| Initial learning rate  | 0.01      |

**Table 3.** Details of the forest fire and smoke dataset.

| Dataset | Training | Validation | Testing |
|---------|----------|------------|---------|
| Number  | 360      | 45         | 45      |

### 3.2. Ablation Experiments

In the ablation experiment, we used the average mAP@0.5 as a criterion for judgment. In addition, the FPS of the model detection process is used to assist in the evaluation. The experimental procedure for each ablation experiment is as follows. First, we set the training

parameters and used the divided forest fire and smoke detection datasets for training. Second, we tested the trained model using a test set of forest fire and smoke datasets to obtain the mAP@0.5. The data from specific ablation experiments are shown in Table 4.

**Table 4.** The data from the specific ablation experiments.

| Model                       | Precision |       |         | Recall |       |         | mAP@0.5 |       |         | FPS |
|-----------------------------|-----------|-------|---------|--------|-------|---------|---------|-------|---------|-----|
|                             | Fire      | Smoke | Average | Fire   | Smoke | Average | Fire    | Smoke | Average |     |
| YOLOv5s                     | 54.8      | 58.8  | 56.8    | 58.8   | 50.8  | 54.8    | 54.3    | 44.0  | 58.1    | 58  |
| YOLOv5s + CA                | 56.6      | 58.8  | 57.7    | 56.5   | 52.9  | 54.7    | 55.3    | 45.5  | 60.0    | 64  |
| YOLOv5s + Bi-FPN            | 56.5      | 59.3  | 57.9    | 55.8   | 53.4  | 54.6    | 55.8    | 44.3  | 59.9    | 63  |
| YOLOv5s + RFB               | 50.6      | 65.4  | 58      | 55.8   | 52.8  | 54.3    | 58.5    | 48.0  | 61.4    | 56  |
| YOLOv5s + CA + Bi-FPN       | 55.5      | 56.1  | 55.8    | 60.5   | 51.3  | 55.9    | 54.8    | 45.7  | 60.8    | 64  |
| YOLOv5s + CA + Bi-FPN + RFB | 61.3      | 59.5  | 60.4    | 56.8   | 48    | 52.4    | 58.8    | 47.3  | 62.2    | 62  |

### 3.3. Comparison

Since we are concerned with the average performance of forest fire and smoke detection, we use the average mAP@0.5 as an evaluation criterion. Experiment 1 shows that YOLOv5 is not very effective in detecting forest fires and smoke. The average mAP@0.5 was only 58.1%.

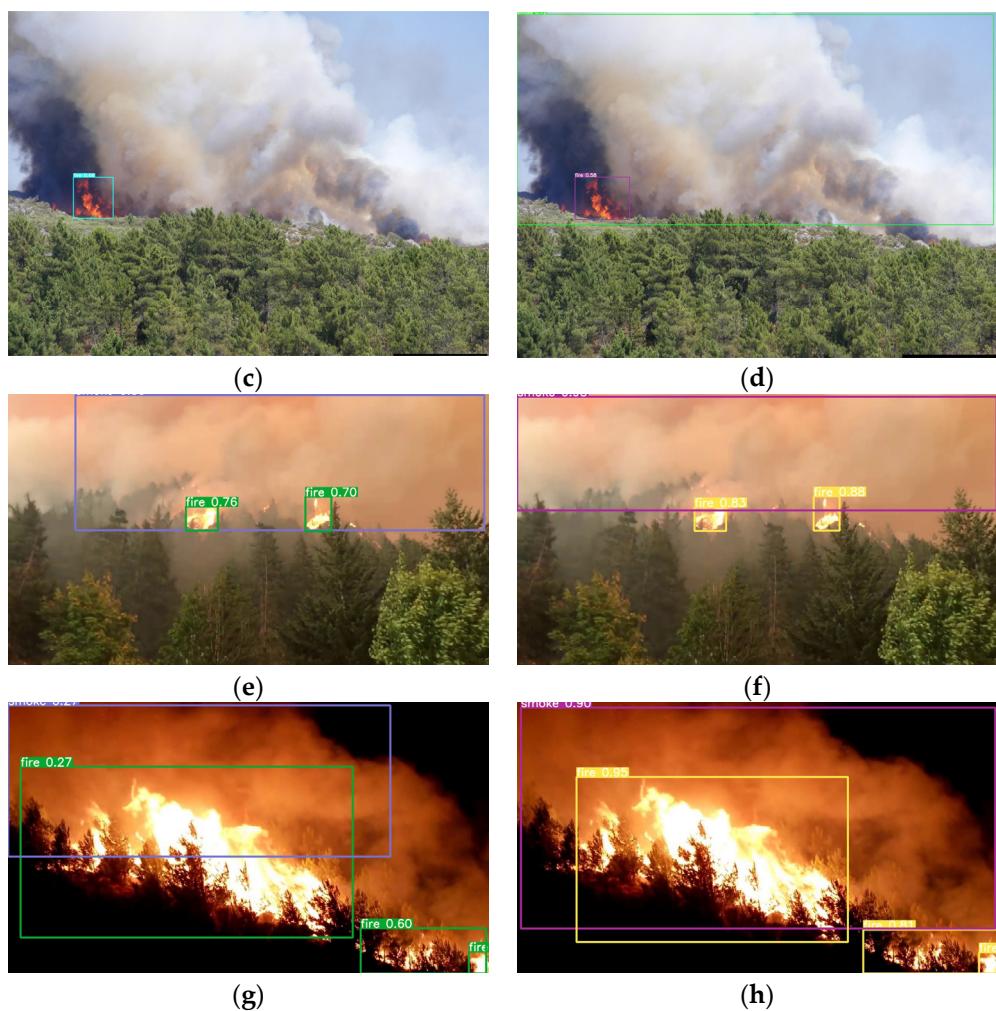
Experiments 2–4 serve to demonstrate the effectiveness of adding a CA module to YOLOv5, replacing the SPPF module with an RFB module, and replacing the PANet structure with Bi-FPN. In Experiment 2, the average mAP@0.5 increases by 1.9% over YOLOv5s. This proves that adding CA modules to YOLOv5s is effective. In Experiment 3, the average mAP@0.5 is 1.8% higher than YOLOv5s, so we can conclude that Bi-FPN can achieve better feature fusion than the PANet structure used in YOLOv5. In Experiment 4, the average mAP@0.5, compared to YOLOv5s, is improved by 3.3%. We can also see that compared to the SPPF module; the RFB module can obtain better global information on forest fire and smoke targets.

Experiments 5 and 6 are to demonstrate that the fusion of the three improved methods mentioned in this paper leads to better results. In experiment 5, the average mAP@0.5, compared to YOLOv5s, is improved by 2.9%, which is higher than both experiment 2 and experiment 3. In Experiment 6, the average mAP@0.5 was significantly higher by 4.1% compared to YOLOv5s and was higher than in both Experiments 4 and 5. This effectively demonstrates the improved forest fire and smoke. Compared with the original YOLOv5s model, our detection model is significantly improved.

The comparison of the detection results between YOLOv5s and our improved forest fire and smoke detection model is shown in Figure 7.



**Figure 7.** Cont.



**Figure 7.** (a,c) are the detection results of the YOLOv5 detection model, neither of which identifies smoke well; (b,d) are the detection results of the improved model, which can accurately identify smoke and forest fires, and the values are also improved; (e,g) are the detection results of YOLOv5 detection model, in the case of both fire and smoke detection without (f,h) after the detailed improvement of the model, and the improved model recognition value improved.

#### 4. Discussion

Forest fires and smoke are diverse in physical form, and different types of targets have different shapes and textures. In addition, compared to ordinary fire smoke, fire smoke in a forest context is more complex, and in addition to the characteristics of the smoke itself, there are a variety of large and small trees and forest fires that can easily affect target detection. Therefore, the detection of forest fires and smoke is more challenging than other detection tasks. Last but not least, it is essential to study an economical and effective forest fire and smoke detection system due to the limited funding of the Chinese forestry sector.

The original model of YOLOv5s cannot effectively focus on various forest fire and smoke targets. To address this issue, we first added a CA module to YOLOv5 to enable better focus on various targets. However, with the development of deep learning, more advanced focus modules will be proposed in the future. Moreover, there will be more appropriate ways to incorporate attention modules on forest fire and smoke detection targets. We will continue to improve our model in subsequent studies.

To better focus on the global information of the target, we replaced the SPPF with the RFB module. The RFB module adds a cavity convolution to Inception, which effectively increases the receptive field. However, replacing the SPPF with the RFB brings a para-

metric boost, and in subsequent studies, we will work on reducing the number of model parameters to give our model better real-time performance.

Bi-FPN not only has a smaller number of parameters but also has a more advanced feature fusion effect than PANet. In the experiments, simply replacing PANet with Bi-FPN also improves the mAP. Although it is only a small parameter reduction, it also gives better real-time performance to the model.

In the follow-up study, we will make further improvements to the forest fire and smoke detection models. First, the current forest fire and smoke detection models still need improvement in smoke identification. We will analyze the reasons for the low smoke recognition rate and make further improvements. Second, our model is based on a Charge-coupled Device (CCD) camera for recognition. We will consider how to select suitable CCDs for forest fire and smoke detection in subsequent studies, and we will also consider how to reduce the impact of complex background detection of forest fire and smoke due to the vulnerability of CCDs to light and shadow. Finally, this forest fire and smoke detection model is still in the laboratory stage, and we will also consider how to deploy this detection model in future studies.

## 5. Conclusions

Smoke from forest fires has a complex background and varies in texture and shape. Traditional target detection models are good for detecting targets such as fire smoke, but they are not effective in detecting forest fires and smoke.

In order to perform effective smoke and fire identification, we did the following work. First, we chose the YOLOv5 model, which is widely used in the field of target detection. Secondly, we made three improvements to the YOLOv5 model due to its poor detection of forest fires and smoke. The CA module was added to make the model better able to focus on forest fire and smoke targets. The SPPF module was replaced with the RFB module to enable the model to focus better on the global information of forest fire smoke targets. Replacing PANet with Bi-FPN enables the model to better perform feature fusion. Finally, we experimentally verified that our model was effectively improved compared to the original YOLOv5s model.

However, the forest fire and smoke detection models proposed in this paper still have some shortcomings. For example, although the CA attention mechanism has made some improvements for all kinds of detection, it is better in smoke recognition, while Bi-FPN has significant improvement effects in forest fire detection. The final model in this paper has enhancements for YOLOv5's original model, but the value of the RFB module, which detects smoke best, has decreased. Therefore, further optimization of this model is considered. Firstly, we will further search for a model that is better for Forest fire detection and a better model for smoke detection for the integration of two types of forest fires and smoke with different distribution and target shapes. Different detection models and attention mechanisms will produce different detection effects. Secondly, further photography or search for images of the early stage of fire is considered to expand the data set and improve its reliability of the data set.

In future work, we will continue to improve the model by seeking more effective and less parameter-intensive methods. We will also investigate methods for deploying forest fire and smoke detection models.

**Author Contributions:** J.L. devised the programs and drafted the initial manuscript and contributed to writing embellishment. R.X. helped with data collection, data analysis. Y.L. 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 National Key R&D Program of China (grant number 2017YFD0600904).

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

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

## References

1. Ghaderpour, E.; Vujadinovic, T. The Potential of the Least-Squares Spectral and Cross-Wavelet Analyses for Near-Real-Time Disturbance Detection within Unequally Spaced Satellite Image Time Series. *Remote Sens.* **2020**, *12*, 2446. [[CrossRef](#)]
2. Jang, E.; Kang, Y.; Im, J.; Lee, D.W.; Yoon, J.; Kim, S.K. Detection and Monitoring of Forest Fires Using Himawari-8 Geostationary Satellite Data in South Korea. *Remote Sens.* **2019**, *11*, 271. [[CrossRef](#)]
3. Marsha, A.L.; Larkin, N.K. Evaluating satellite fire detection products and an ensemble approach for estimating burned area in the United States. *Fire* **2022**, *5*, 147. [[CrossRef](#)]
4. Duan, F.; Wang, Y.N.; Lei, X.F. A review of machine vision technology and its applications. *Autom. Expo* **2002**, *19*, 59–61.
5. Dai, L. *Research and Application of Algorithm for Smoke and Fire Detection in Intelligent Monitoring System*; Beijing University of Posts and Telecommunications: Beijing, China, 2015.
6. Genovese, A.; Labati, R.D.; Piuri, V. Wildfire smoke detection using computational intelligence techniques. In Proceedings of the IEEE International Conference on Computational Intelligence for Measurement Systems Applications, Ottawa, ON, Canada, 19–21 September 2011.
7. Yuan, F. Video-Based smoke detection with histogram sequence of lbp and lbpv pyramids. *Fire Saf. J.* **2011**, *46*, 132–139. [[CrossRef](#)]
8. Xue, Q.; Lin, H.; Wang, F. FCDM: An Improved Forest Fire Classification and Detection Model Based on YOLOv5. *Forests* **2022**, *13*, 2129. [[CrossRef](#)]
9. Qian, J.; Lin, H. A Forest Fire Identification System Based on Weighted Fusion Algorithm. *Forests* **2022**, *13*, 1301. [[CrossRef](#)]
10. Xue, Z.; Lin, H.; Wang, F. A Small Target Forest Fire Detection Model Based on YOLOv5 Improvement. *Forests* **2022**, *13*, 1332. [[CrossRef](#)]
11. Xue, X.; Yuan, F.; Zhang, L.; Yang, L.; Shi, J. From traditional to deep: Visual smoke recognition, detection and segmentation. *Chin. J. Graph.* **2019**, *24*, 1627–1647.
12. Lin, J.; Lin, H.; Wang, F. STPM\_SAHI: A Small-Target Forest Fire Detection Model Based on Swin Transformer and Slicing Aided Hyper Inference. *Forests* **2022**, *13*, 1603. [[CrossRef](#)]
13. Lin, J.; Lin, H.; Wang, F. A Semi-Supervised Method for Real-Time Forest Fire Detection Algorithm Based on Adaptively Spatial Feature Fusion. *Forests* **2023**, *14*, 361. [[CrossRef](#)]
14. Chen, G.; Zhou, H.; Li, Z.; Gao, Y.; Bai, D.; Xu, R.; Lin, H. Multi-Scale Forest Fire Recognition Model Based on Improved YOLOv5s. *Forests* **2023**, *14*, 315. [[CrossRef](#)]
15. Bakri, N.S.; Adnan, R.; Ruslan, F.A. A methodology for fire detection using colour pixel classification. In Proceedings of the 2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA), Penang, Malaysia, 9–10 March 2018; IEEE: Batu Ferringhi, Malaysia, 2018; pp. 94–98.
16. Wang, T.; Bu, L.; Zhou, Q.; Yang, Z. A new fire recognition model based on the dispersion of color component. In Proceedings of the 2015 IEEE International Conference on Progress in Informatics and Computing (PIC), Nanjing, China, 18–20 December 2015; IEEE: Nanjing, China, 2015; pp. 138–141.
17. 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 Futuristic Trends in Research & Innovation for Social Welfare, Coimbatore, India, 29 February–1 March 2016; IEEE: Coimbatore, India, 2016; pp. 1–6.
18. Wang, T.; Shi, L.; Yuan, P.; Bu, L.; Hou, X. A new fire detection method based on flame color dispersion and similarity in consecutive frames. In Proceedings of the Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; IEEE: Jinan, China, 2018; pp. 151–156.
19. Lei, S.; Fangfei, S.; Teng, W.; Leping, B.; Xinguo, H. A new fire detection method based on the centroid variety of consecutive frames. In Proceedings of the 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, 2–4 June 2017; IEEE: Chengdu, China, 2017; pp. 437–442.
20. Gunawaardena, A.E.; Ruwanthika, R.M.M.; Jayasekara, A.G.B.P. Computer vision based fire alarming system. In Proceedings of the 2016 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 5–6 April 2016; IEEE: Moratuwa, Sri Lanka, 2016; pp. 325–330.
21. Ultralytics-YOLOv5. Available online: <https://github.com/ultralytics/YOLOv5> (accessed on 5 June 2022).
22. Yuan, Z.; Liu, Z.; Zhu, C.; Qi, J.; Zhao, D. Object Detection in Remote Sensing Images via Multi-Feature Pyramid Network with Receptive Field Block. *Remote Sens.* **2021**, *13*, 862. [[CrossRef](#)]
23. 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.
24. Ju, C.; Guan, C. *Tensor-CSPNet: A Novel Geometric Deep Learning Framework for Motor Imagery Classification*; IEEE: New York, NY, USA, 2022.

25. Wang, C.Y.; Liao HY, M.; 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 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; IEEE: New York, NY, USA, 2020.
26. Chatzipantazis, E.; Pertigkiozoglou, S.; Dobriban, E.; Daniilidis, K. SE(3)-Equivariant Attention Networks for Shape Reconstruction in Function Space. *arXiv* **2022**, arXiv:2204.02394.
27. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
28. Liu, S.; Huang, D. Receptive field block net for accurate and fast object detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 385–400.
29. Lin, H.; Han, Y.; Cai, W.; Jin, B. Traffic Signal Optimization Based on Fuzzy Control and Differential Evolution Algorithm. In *IEEE Transactions on Intelligent Transportation Systems*; IEEE: New York, NY, USA, 2022. [[CrossRef](#)]
30. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
31. Zheng, Z.; Lei, L.; Sun, H.; Kuang, G. FAGNet: Multi-Scale Object Detection Method in Remote Sensing Images by Combining MAFPN and GVR. *J. Comput.-Aided Des. Comput. Graph.* **2021**, *33*, 883–894. [[CrossRef](#)]
32. Lin, H.; Tang, C. Analysis and Optimization of Urban Public Transport Lines Based on Multiobjective Adaptive Particle Swarm Optimization. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 16786–16798. [[CrossRef](#)]
33. 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.* **2022**, *23*, 9869–9882. [[CrossRef](#)]
34. Zhang, L.; Yuan, F.; Zhang, W.; Zeng, X. A review of research on fully convolutional neural networks. *Comput. Eng. Appl.* **2020**, *56*, 25–37.

**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.