



# Article An Efficient Forest Fire Detection Algorithm Using Improved YOLOv5

Pei Shi<sup>1,2,\*</sup>, Jun Lu<sup>1</sup>, Quan Wang<sup>2</sup>, Yonghong Zhang<sup>1,2</sup>, Liang Kuang<sup>3</sup> and Xi Kan<sup>2</sup>

- <sup>1</sup> School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China; lu6288128@163.com (J.L.)
- <sup>2</sup> School of IoT Engineering, Wuxi University, Wuxi 214105, China; seepwq@163.cn (Q.W.); kanxi@cwxu.edu.cn (X.K.)
- <sup>3</sup> School of IoT Engineering, Jiangsu Vocational College of Information Technology, Wuxi 214153, China; kuangl@jsit.edu.cn
- \* Correspondence: ship@cwxu.edu.cn

Abstract: Forest fires result in severe disaster, causing significant ecological damage and substantial economic losses. Flames and smoke represent the predominant characteristics of forest fires. However, these flames and smoke often exhibit irregular shapes, rendering them susceptible to erroneous positive or negative identifications, consequently compromising the overall performance of detection systems. To enhance the average precision and recall rates of detection, this paper introduces an enhanced iteration of the You Only Look Once version 5 (YOLOv5) algorithm. This advanced algorithm aims to achieve more effective fire detection. First, we use Switchable Atrous Convolution (SAC) in the backbone network of the traditional YOLOv5 to enhance the capture of a larger receptive field. Then, we introduce Polarized Self-Attention (PSA) to improve the modeling of long-range dependencies. Finally, we incorporate Soft Non-Maximum Suppression (Soft-NMS) to address issues related to missed detections and repeated detections of flames and smoke by the algorithm. Among the plethora of models explored, our proposed algorithm achieves a 2.0% improvement in mean Average Precision@0.5 (mAP<sup>50</sup>) and a 3.1% enhancement in Recall when compared with the YOLOv5 algorithm. The integration of SAC, PSA, and Soft-NMS significantly enhances the precision and efficiency of the detection algorithm. Moreover, the comprehensive algorithm proposed here can identify and detect key changes in various monitoring scenarios.

check for **updates** 

**Citation:** Shi, P.; Lu, J.; Wang, Q.; Zhang, Y.; Kuang, L.; Kan, X. An Efficient Forest Fire Detection Algorithm Using Improved YOLOv5. *Forests* **2023**, *14*, 2440. https:// doi.org/10.3390/f14122440

Academic Editor: Costantino Sirca

Received: 17 November 2023 Revised: 3 December 2023 Accepted: 9 December 2023 Published: 14 December 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/). **Keywords:** forest fires; fire detection; YOLOv5; switchable atrous convolution; polarized self-attention; soft non-maximum suppression

## 1. Introduction

To this date, multiple large-scale fire incidents have occurred globally. Notable instances of wildfire disasters in 2015, for example, resulted in a tragic loss of 494,000 lives and incurred a substantial financial toll amounting to CNY 22.098 billion [1]. In addition to forest fires, urban fires are a frequent occurrence, often resulting in severe casualties and property losses. In 2019, a fire and explosion tragically transpired in Xiangshui, Jiangsu province, China, claiming a total of 78 lives and causing injuries to 617 individuals across the province [2]. In 2015, a major fire and explosion rocked Tianjin Port in China, directly causing 165 fatalities and an estimated loss of CNY 6.866 billion [3]. It is evident that fires pose a significant threat. Compared to urban fires, forest fires have a larger burning area. Moreover, the abundance of combustible materials in forests leads to rapid fire propagation. Additionally, the complex terrain of forests makes firefighting efforts extremely challenging. Therefore, the aim of this paper is to accurately detect forest fires for an effective response to them.

Traditional fire detection techniques rely on sensor-based detection, including smoke detection, temperature detection, photoelectric detection, gas detection, flame detection, etc. The perception data are gathered through temperature sensors, smoke detectors, carbon

monoxide sensors, and various other sensing devices. Subsequently, neural networks are employed to make predictions based on the acquired data [4]. This sensor-based detection method leverages various physical parameters from the field environment, providing comprehensive and detailed fire monitoring data, thus aiding in forecasting fire development trends. Furthermore, federated learning based on distributed machine learning can handle one-dimensional fire data [5]. Solórzano et al. [6] conducted research focused on optimizing the performance of gas sensor arrays. Their work demonstrated that systems based on gas sensor arrays can offer faster fire alarm responses when compared to conventional smokebased fire detectors. In another study, Sun et al. [7] developed an automated enhanced multi-trend back-propagation neural network algorithm. This algorithm utilizes multiple neural networks to enhance predictions and adds a special trend extraction component. This component helps the algorithm to detect fire initiation, even with small temperature changes. Pang et al. [8] used a random forest [9] model to predict the probability of forest fires in China. Based on these probabilities, they generated a forest fire occurrence probability map for China and a seasonal forest fire probability map. Kalantar et al. [10] used three machine learning models and 14 sets of fire predictors derived from vegetation indices, climatic variables, environmental factors, and topographical features to predict forest fire susceptibility in the Chaloos Rood watershed in Iran. However, the accuracy of sensorbased detection is contingent upon equipment maintenance, which can pose a challenge in scenarios with limited budgets. Moreover, where sensors are usually positioned in open environments such as airports, stadiums, outdoor venues, and instances away from the actual fire source, it becomes challenging to collect information like gas concentrations and temperature. Consequently, the efficiency and accuracy of sensor data collection tend to be relatively low. To alleviate these issues, a vision-based approach has been employed to facilitate real-time fire detection.

Vision-based object detection systems have the ability to analyze image or video data for fire detection without physical contact with the target object. These systems can be effectively used in various fire detection scenarios. Notably, deep learning-based methods have demonstrated robust learning capabilities and scalability [11], rendering them widely applicable in the field of fire detection. With the continuous advancements in deep learning and convolutional neural networks [12], vision-based object detection techniques have achieved remarkable accuracy. These methods can classify and locate targets with high precision, enabling accurate object detection and recognition [13]. Zhao et al. [14] proposed a forest fire and smoke recognition method based on the fusion of environmental information, using Faster R-CNN [15] as the fundamental framework. Nguyen et al. [16] developed a real-time fire detection system tailored for large-scale monitoring, utilizing unmanned aerial vehicles (UAVs) equipped with an integrated visual detection and alarm system. Their approach was grounded in the Single Shot MultiBox Detector (SSD) [17]. Zheng et al. [18] employed the principal component analysis (PCA) reconstruction technique in conjunction with the dynamic convolutional neural network (DCNN) [19]. This approach led to the development of a 15-layer DCNN model capable of accurately identifying critical fire-prone areas. Li et al. [20] addressed the challenge of poor small object detection within the DEtection TRansformer (DETR) [21] by combining a lightweight normalized attention module with DETR when detecting fires. With technology continually advancing and computing power improving, the YOLO series of algorithms have gained prominence. These algorithms strike a favorable balance between accuracy and detection speed, and researchers have progressively begun applying them to fire detection tasks. Qin et al. [22] introduced a fire classification method that combines depthwise separable convolution with YOLOv3 [23]. They use depthwise separable convolution to classify fire images, and then apply YOLOv3 to locate fire in the images. This approach improves the detection accuracy and efficiency. Wang et al. [24] introduced a model named Light-YOLOv4, which employs MobileNetv3 [25] as the backbone network instead of YOLOv4 [26]. They seamlessly integrated the bidirectional feature pyramid network (BiFPN) [27] and the depthwise separable attention module into YOLOv4, leading to a substantial reduction in parameters while maintaining nearly the same level of accuracy as

the original model. This innovative approach achieves a lightweight effect. Bahhar et al. [28] combined a YOLO architecture with two weights with a voting ensemble CNN architecture. If the CNN detects anomalies in a frame, the YOLO architecture is then utilized to locate the smoke or fire. Yu et al. [29] made improvements to the image preprocessing, modules, and loss functions of YOLOv5 to achieve better detection performance. Dou et al. [30] introduced the Convolutional Block Attention Module (CBAM) [31], BiFPN, and transposed convolution into YOLOv5, significantly improving the algorithm's accuracy in fire detection. Du et al. [32] introduced ConvNeXt [33] into YOLOv7 [34], resulting in the development of a fire detection model that effectively balances detection speed and accuracy. Chen et al. [35] utilized Ghost Shuffle Convolution [36] to construct two lightweight modules, incorporating these into YOLOv7. This effectively reduced the model's parameter count and accelerated the model's convergence speed. Additionally, they integrated Coordinate Attention [37] into the YOLOv7 feature extraction network to enhance smoke feature extraction and attenuate background interference. Talaat et al. [38] proposed an improved method for smart city fire detection based on the YOLOv8 algorithm. It leverages the advantages of deep learning to detect specific features of fires in real time. Some classic two-stage detection algorithms and the DETR model have high accuracy, but they are slow and large. These limitations prevent real-time fire detection YOLO algorithms from balancing detection accuracy and speed well. The YOLO series has been updated with multiple versions, with various modules in YOLO optimized multiple times and continuously incorporating advanced optimization strategies. After multiple versions of optimization, YOLOv5 addresses some drawbacks of its predecessors, such as large parameter count, and achieves improved detection performance. Many scholars in fire detection use YOLOv5 as one of the main algorithms. Therefore, this paper uses a method based on YOLOv5 for fire detection. However, there is still room for improvement in YOLOv5. For instance, YOLOv5 tends to miss smoke detection in the context of forest fires. Moreover, when smoke and flames are densely distributed, occurrences of missed detections also arise. This paper aims to improve YOLOv5 to meet the requirements of fire detection.

In this paper, we propose an improved detection algorithm based on YOLOv5, named SPS-YOLOv5, to enhance the detection accuracy and reduce the false detection rates of forest fires. Firstly, to address the issue regarding the confusion of flames, smoke, and background information in widespread distribution scenes, we adopt the SAC module [39]. SAC can expand the convolution's receptive field, allowing for the comprehensive extraction of background information to differentiate it from the detected objects. Then, to tackle the problem of insufficient feature extraction for flames and smoke, the PSA [40] attention mechanism is introduced into the neck network. This attention mechanism captures global long-range dependencies, thereby enhancing the feature extraction capability of convolution. Finally, to address the problem of missed detections caused by multiple targets overlapping, Soft-NMS [41] is deployed to replace the original NMS [42] to capture more useful information and achieve better detection results.

The remaining sections of this paper are organized as follows. Section 2 provides a detailed introduction to the improved YOLOv5 algorithm proposed in this study. In Section 3, we present the dataset used, the experimental environment, parameter settings, and evaluation metrics. Sections 4 and 5 showcase the experimental results, followed by a discussion and analysis of the results and the proposed algorithm. Additionally, we outline future work prospects. In Section 6, a summary of the work presented in this paper is provided.

#### 2. Methods

## 2.1. YOLOv5m Algorithm

YOLOv5 is a one-stage object detection algorithm that uses a backbone network for feature extraction. It detects objects of different sizes with multi-level feature fusion and multi-scale prediction. The YOLOv5m algorithm has three main components: the backbone, neck, and head, as shown in Figure 1.

- Backbone: In the YOLOv5m model, CSPDarknet53 serves as the backbone network. CSPDarknet53 is rooted in the Darknet53 network architecture and incorporates Cross Stage Partial (CSP) [43] connections. These connections enhance the network's ability for extracting image features and computational efficiency;
- Neck: This section employs a combined architecture that consists of the Feature Pyramid Network (FPN) [44] and the Path Aggregation Network (PAN) [45]. FPN propagates semantic information top-down, while PAN conveys localization information top-down. Then, utilizing both up-sampling and down-sampling operations, it effectively integrates feature maps from various levels to produce a multi-scale feature pyramid. This pyramid enhances the algorithm's capacity to capture information at various scales;
- Head: This component consists of four parts: Anchors, Convolutional Layers, Prediction Layers, and Non-Maximum Suppression. The anchors are a predefined set of bounding boxes used to generate candidate boxes on the feature map. The convolutional layers in the detection head are responsible for processing the feature maps and extracting features. Each prediction layer is responsible for predicting a set of bounding boxes and their corresponding class probabilities. In the output bounding boxes, the non-maximum suppression algorithm is employed to suppress overlapping boxes, retaining only the most representative ones. The primary function of the head is to extract features and make predictions for object detection. It extracts high-level features from the input image and utilizes them to predict the position and category of objects in the image.



Figure 1. The framework of YOLOv5m network.

## 2.2. Switchable Atrous Convolution

Atrous convolution [46] introduces a dilation rate in conventional convolution, thereby enabling each convolution output to cover a larger range of information. However, this enlarges the receptive field and may cause the loss of local information continuity. The SAC module alleviates this issue by utilizing the ability to switch between ordinary convolution and atrous convolution. This feature alleviates the problem of discontinuous information extraction caused by atrous convolution while simultaneously expanding the receptive field.

Figure 2 shows the overall structure of the SAC. It mainly consists of three parts: two global modules located before and after the SAC module, and the SAC module. Let y = Conv(x, w, r) be represented as a convolution operation with weight w and dilation rate r. This convolution operation takes x as input and produces y as output. The conversion of a regular convolution to SAC can be expressed as follows:

$$Conv(\mathbf{x}, \mathbf{w}, 1) \xrightarrow[to SAC]{Convert} S(\mathbf{x}) \cdot Conv(\mathbf{x}, \mathbf{w}, 1) + (1 - S(\mathbf{x})) \cdot Conv(\mathbf{x}, \mathbf{w} + \Delta \mathbf{w}, r)$$
(1)

where *x* represents the input, *w* and  $\Delta w$  represent the weight parameters, *w* is a pre-trained checkpoint, and  $\Delta w$  is the weight initialized to 0 at the same scale as *w*. *r* represents the dilation rate of the convolution operation. *S*(•) performs 5 × 5 average pooling and 1 × 1 convolution as its operations. The functionality of the switch depends on both the input and its position. As a result, regular convolution and atrous convolution can be fused to achieve more efficient feature extraction.



Figure 2. Architecture diagram of SAC module.

The SAC module introduces a locking mechanism, wherein the weights of the first convolution in Equation (1) are set to w, and the weights of the second convolution are set to  $w + \Delta w$ . The two lightweight global context modules, before and after the SAC module, work as follows: They first employ global average pooling on the input features to compress the information and complete convolution. The output is directly integrated back into the primary stream without any extra complex operations.

## 2.3. Polarized Self-Attention

The attention mechanism helps the model focus on important regions or features, which improves the localization and recognition of objects in an image. The combination of convolutional fusion with attention mechanisms amplifies the feature extraction ability of convolutions, ultimately enhancing the accuracy of object detection. The proposed method employs Polarized Self-Attention (PSA) to capture long-range dependencies and improve the convolutional feature extraction capability.

The main idea of PSA is to compress features along one direction while increasing the intensity range of the loss. This attention mechanism has two distinct components: (1) Filtering: This component collapses one dimension of features completely while keeping a high resolution in orthogonal dimensions. (2) High Dynamic Range (HDR): First, in the attention module, the Softmax function is applied to the smallest tensor to extend the attention range. Then, the Sigmoid function is employed for dynamic mapping. A detailed depiction of the PSA structure is presented in Figure 3.



Figure 3. The architecture diagram of PSA module.

From Figure 3, it can be observed that, in the channel branch, this architecture uses a  $1 \times 1$  convolution to transform the input feature X into Q and V. In this transformation, the channel dimension of Q is entirely compressed, while the channel dimension of V is retained at a relatively high level. Since the channel dimension of Q is compressed, the Softmax function is used to enhance the information of Q. Then, matrix multiplication is performed between Q and V, followed by applying a  $1 \times 1$  convolution and Layer Normalization (LN) to increase the channel dimension from C/2 to C. Finally, the Sigmoid function is applied to ensure that all parameters are constrained within the range from 0 to 1 in the channel branch. In the spatial branch, a similar process is followed. The input features are transformed into Q and V using a  $1 \times 1$  convolution. It uses Global Pooling on Q to reduce the spatial dimension to  $1 \times 1$ . However, the spatial dimension of the V is high. Given the compression of the spatial dimension of Q, the Softmax function is applied to enhance the information of Q. the spatial branch is applied to enhance the information of the spatial dimension of the spatial dimension of the spatial dimension of the Softmax function is applied to enhance the information of 2 and 3. The spatial dimension of 3 and 3 a

#### 2.4. Soft-NMS

Non-Maximum Suppression (NMS) is a widely employed computer vision algorithm, commonly utilized in tasks such as object detection and bounding box regression. The primary objective of NMS is to tackle the problem of overlapping candidate bounding boxes by eliminating redundant bounding boxes.

The working principle of NMS is as follows: Given a set of candidate bounding boxes, NMS starts by choosing the bounding box with the highest confidence score. It continues to calculate the degree of overlap, typically using Intersection over Union (IoU), between this selected bounding box and all other bounding boxes. Any bounding box with an IoU value with the chosen bounding box exceeding a predefined threshold is then removed. This process iterates by selecting the subsequent bounding box with the highest confidence score and eliminating any bounding boxes with significant IoU overlap. This procedure continues until all bounding boxes have been processed. The NMS algorithm can be mathematically expressed as follows:

$$s_i = \begin{cases} s_i, & iou(M, b_i) < N_t \\ 0, & iou(M, b_i) \ge N_t \end{cases}$$
(2)

where  $b_i$  represents the *i*th detection bounding box,  $S_i$  is the corresponding detection score,  $N_t$  is the NMS threshold, M stands for the detection bounding box with the highest detection score, and IoU denotes the Intersection over Union, which is calculated as the ratio of the intersection area of two bounding boxes to their union area.

This NMS adopts a binary method, which removes bounding boxes that overlap with the one with the highest confidence score. However, this may eliminate some bounding boxes that have lower confidence scores but overlap a lot with the target object and affect the object detection performance. Soft-NMS solves this issue by introducing a decay function that gradually diminishes the confidence scores of other bounding boxes based on their overlap with the one possessing the highest confidence score, rather than immediately discarding them. This technique keeps some information from overlapping bounding boxes. There are two decay functions: linear and Gaussian, which can be expressed as follows:

$$s_i = \begin{cases} s_i, & iou(M, b_i) < N_t \\ s_i(1 - iou(M, b_i)), & iou(M, b_i) \ge N_t \end{cases}$$
(3)

$$s_i = s_i e^{-\frac{iou(M,b_i)^2}{\sigma}}, \forall b_i \notin D$$
(4)

where *D* represents the set of final detection bounding boxes. Equation (3) represents a linear function, and Equation (4) represents a Gaussian function. It can be observed from Equation (3) that this function is discontinuous, which may cause a score gap in the box set. Therefore, this paper adopts the Gaussian function as the decay function.

#### 2.5. The Proposed Algorithm for Smoke and Fire Detection

During a forest fire, because the distribution of flames and smoke is relatively dense, YOLOv5 may experience missed detections. It also needs to improve the average precision of YOLOv5 in detecting smoke and flames. To capture the characteristics of flame and smoke efficiently and improve the detection accuracy and efficiency of terminal equipment, this paper introduces an enhanced YOLOv5 algorithm, called SPS-YOLOv5 (incorporating SAC, PSA, and Soft-NMS modules into the YOLOv5m framework). The framework of the SPS-YOLOv5 algorithm is depicted in Figure 4.



**Figure 4.** The framework of the SPS-YOLOv5 algorithm: the modules with bold black borders in the picture are improved modules.

In Figure 4, it is evident that we focus on the improvements in the backbone and neck modules. First, we replaced the last four convolutions in the backbone with SAC. Then, SAC is added to the last two Cross-Stage Partial (CSP) modules in the backbone. This makes the backbone network use atrous convolutions better, which provides a larger receptive field and less loss of local information. In the neck section, PSA is introduced into the four convolutions, facilitating the model in capturing long-range dependencies more effectively, making the proposed algorithm more efficient. In the NMS, Soft-NMS replaces the traditional NMS. This helps the algorithm detect and recognize weaker objects better, which improves the diversity and stability of the object detection results.

## 3. Experiment

## 3.1. Datasets

To validate the effectiveness of the detection algorithm, we reconstructed a composite large dataset from the Internet. The dataset is composed of a total of 7878 image samples, encompassing both fire and smoke instances. This dataset consists of mostly forest fire images, in addition to some urban fire images, spanning from small-scale to large-scale incidents. This diverse composition helps improve the algorithm's capacity to generalize across various fire-related situations. Some of the forest fire images in this dataset are from drone photography. The dataset was divided into training, validation, and test sets in a ratio of 7:1:2. Consequently, the training set comprises 5514 images, the validation set comprises 788 images, and the test set has 1576 images. A selection of sample images from the dataset is presented in Figure 5 for reference.





Figure 5. Sample images of the self-created dataset: (a) Small-scale fire; (b) Small-scale fire; (c) Large-scale fire; (d) Large-scale fire.

#### 3.2. Experiment Setup

In this study, we use PyTorch to experiment, and the hardware details are shown in Table 1. The PyTorch version is 1.13.1, and the Python version is 3.9. The graphics card model is the GTX 1080 Ti.

Item	Туре				
CPU	Intel(R) Core(TM) i7-12700H 2.70 GHz				
GPU	NVIDIA GeForce GTX 1080 Ti 11G				
Programming language	Python 3.9				
Operating system	Windows 11				
Deep learning framework	PyTorch 1.13.1				

Table 1. Description of experimental conditions.

The experiment employed the SGD optimizer with pre-trained weights from YOLOv5m for training. As illustrated in Figure 6, the mean Average Precision (mAP)@0.5 value reached its peak and stabilized at approximately the 70th epoch. To mitigate overfitting, the experiment was limited to 70 epochs by experience. The batch size was 8, and a momentum value of 0.937 was applied. The images had a fixed aspect ratio of 1:1, with dimensions of  $640 \times 640$  pixels. For fire detection, the algorithm adopted an anchorbased approach, utilizing anchor boxes generated through the anchor mechanism. These anchor boxes served as prior bounding boxes, enabling the network to conduct object classification and bounding box regression directly. This approach facilitated faster convergence and improved the training consistency of the algorithm.



Figure 6. mAP<sup>50</sup> curve of SPS-YOLOv5.

## 3.3. Model Evaluation Metrics

To assess the detection performance of fire and smoke, this study utilizes the following evaluation metrics: mean Average Precision (mAP), Recall (R), Frames Per Second (FPS), parameters, and Floating Point Operations (FLOPs). FPS is employed to quantify the real-time detection capability of the algorithm, while parameters are used to evaluate the algorithm's complexity and capacity. FLOPs show the computational load or complexity of the algorithm. Higher values of mAP, Recall, and FPS indicate superior algorithm performance, while lower values of parameters and FLOPs imply a more concise algorithm with enhanced computational efficiency. The formulas for calculating mAP and Recall are as follows:

$$Precision = \frac{TP}{TP + FP}$$
(5)

$$\operatorname{Recall} = \frac{\operatorname{TP}}{\operatorname{TP} + \operatorname{FN}} \tag{6}$$

$$AP = \int_0^1 P(r)dr$$
(7)

$$mAP = \frac{\sum_{i=1}^{C} AP_i}{C}$$
(8)

where TP, FP, and FN represent the numbers of true positives, false positives, and false negatives, respectively. *C* stands for the number of classes in the dataset, and in this study, the dataset covers two classes of objects, fire and smoke, so C = 2.

## 4. Results

## 4.1. YOLO Series Algorithms Comparison

YOLO is a series of algorithms widely employed for object detection. It makes the object detection a regression problem that can be executed in a single forward pass. It predicts bounding boxes and class labels within grid cells in the image. The algorithm has significantly improved from the first version to the latest one. In order to compare the detection performance of different YOLO series on this dataset, we present the detection results in Table 2, and the mAP<sup>50</sup> and Recall results in Figure 7.

Table 2. A comparative analysis of various versions of YOLO in terms of different indices.

Number	Method	mAP <sup>50</sup> (%)	R (%)	FPS	Parameters (M)	FLOPs (G)
1	YOLOv3	77.4	58.8	56	61.5	65.6
2	YOLOv4	65.3	38.2	45	63.9	60.0
3	YOLOv5	77.8	71.2	65	20.1	47.9
4	YOLOv7	77.3	72.8	68	36.5	103.2
5	YOLOv8	77.6	74.0	45	25.8	78.7



Figure 7. YOLO series performance comparison.

From Table 2, it can be clearly seen that YOLOv5 achieved a mAP<sup>50</sup> of 77.8%, better than other YOLO versions in terms of detection precision. Even though YOLOv5's FPS is slightly lower than YOLOv7 at 65, it still outperforms YOLOv3, YOLOv4, and YOLOv8 in processing speed. This makes YOLOv5 a good choice for real-time or high-speed fire monitoring applications. Although the Recall value of YOLOv5 is lower than that of YOLOv8, it is better than YOLOv8 in mAP. And because YOLOv8 has a more complex

network architecture, including multiple residual units and multiple branches, its FPS, parameters, and FLOPs are all higher than for YOLOv5. Additionally, YOLOv5 has much fewer parameters (20.1 M) and lower FLOPs (47.9 G) than its YOLO counterparts in the comparison, indicating a lower algorithm complexity. In summary, compared to other YOLO versions, YOLOv5 has superior detection precision, faster processing speed, and lower algorithm complexity for fire detection. Consequently, this paper adopts an improved YOLOv5-based fire detection algorithm.

## 4.2. Analysis of Ablation Experiments

To evaluate the performance of the proposed method, we conducted ablation experiments on SPS-YOLOv5. We introduced the improved modules to the YOLOv5 baseline, and then constructed the YOLOv5 + SAC and YOLOv5 + SAC + Soft-NMS algorithms. To compare the algorithms' performance for different target sizes and difficulty levels, we calculated mAP values at two IoU thresholds: 0.5 and 0.75. The experimental results are summarized in Table 3, and Figure 8 shows the loss curves of these algorithms in the ablation process.

Model	mAP <sup>50</sup> (%)	mAP <sup>75</sup> (%)	R (%)	FLOPs (G)	FPS
YOLOv5	77.8	54.0	71.2	47.9	65
YOLOv5 + SAC	78.5	53.1	73.7	30.9	47
YOLOv5 + SAC + Soft-NMS	78.8	58.9	73.7	30.9	45
SPS-YOLOv5	79.8	58.3	74.3	31.7	43



Table 3. Detection results of ablation experiment.

**Figure 8.** Loss curves during ablation experiments: (**a**) Loss curve of YOLOv5; (**b**) Loss curve of YOLOv5 + SAC; (**c**) Loss curve of SPS-YOLOv5.

From Table 3, it is evident that the inclusion of the SAC module significantly enhances mAP<sup>50</sup> and reduces the computational complexity of the algorithm. This indicates that, while improving mAP<sup>50</sup>, the overall complexity of the algorithm decreases. Furthermore, when using YOLOv5 + SAC + Soft-NMS, it improves mAP<sup>50</sup> slightly and mAP<sup>75</sup> significantly, without changing the algorithm's computational complexity. In summary, the detection results of SPS-YOLOv5 are much better than those of the original YOLOv5. It has a 2.0% increase in mAP<sup>50</sup>, a 4.3% increase in mAP<sup>75</sup>, a 3.1% increase in Recall, and a significant reduction in FLOPs. This implies that, in comparison to the original algorithm, the improved algorithm has superior detection performance for fire and smoke, while also decreasing the algorithm's computational complexity. This makes it more suitable for training and inference.

Figure 8 shows the loss curves of different algorithms during training. We can see that adding the SAC module makes all three algorithms have better loss curves with less oscillations. Because Soft-NMS is only applied for validation and prediction, YOLOv5 + SAC + Soft-NMS has the same loss curve as YOLOv5 + SAC. For the SPS-YOLOv5 algorithm, which incorporates all three modules, it displays a smoother object loss compared to the algorithm by just adding the SAC module. SPS-YOLOv5 converges faster and smoother than YOLOv5. This means that the improved algorithm's training is more stable and better. Figure 9 shows the visualization results of detection. The comprehensive visualization results indicate that the proposed SPS-YOLOv5 algorithm has higher confidence scores for the smoke label. It also solves the issue of missed detection. In short, the SPS-YOLOv5 algorithm outperforms the YOLOv5 algorithm in detecting fire and smoke. It is more stable and better, as shown by its loss curves and detection results.



Figure 9. Cont.



**Figure 9.** Detection results of YOLOv5 and SPS-YOLOv5: (**a**) Partial detection results of YOLOv5; (**b**) Partial detection results of SPS-YOLOv5.

## 4.3. Comparison of Different Detection Algorithms

In this subsection, we compared SPS-YOLOv5 with several classical object detection algorithms. The comparison results are displayed in Table 4.

Number	Method	mAP <sup>50</sup> (%)	R (%)	FPS	Parameters (M)	FLOPs (G)
1	SSD	71.8	62.7	78	23.7	60.9
2	Faster R-CNN	75.8	84.9	11	40.3	269.7
3	EfficientDet	78.0	63.3	22	3.8	4.7
4	YOLOv3-EfficientNet	74.0	53.3	58	7.0	3.8
5	CenterNet	75.3	40.7	59	32.7	70.2
6	SPS-YOLOv5	79.8	74.3	43	31.8	31.7

Table 4. Algorithm comparison results.

From Table 4, it is evident that SPS-YOLOv5 achieves the highest mAP<sup>50</sup> at 79.8%. SSD [17] is a one-stage object detection algorithm. It does not require a region proposal stage and can directly generate the class probability and position coordinates of objects. It can directly obtain the final detection results after one detection. When compared to SSD, SPS-YOLOv5 exhibits an impressive 8.0% improvement in mAP<sup>50</sup> and an 11.6% increase in Recall. Additionally, it also has fewer FLOPs. Faster R-CNN [15] is a two-stage detector that conducts region proposal and region classification separately. In comparison to Faster R-CNN, SPS-YOLOv5 has a 4.0% improvement in mAP<sup>50</sup>. Although its Recall is slightly lower than that of Faster R-CNN, it has fewer parameters and FLOPs and higher FPS than Faster R-CNN. Compared to CenterNet [47], which predicts objects as points, SPS-YOLOv5 exhibits a 4.5% improvement in mAP<sup>50</sup> and significantly higher Recall than CenterNet. Furthermore, it has fewer parameters and boasts a remarkable 54.8% reduction in FLOPs.

EfficientDet [27] is also a one-stage detector like SSD. It is improved and extended based on EfficientNet. It has the best performance among other algorithms we compared it with. However, a detailed comparison between EfficientDet and SPS-YOLOv5 reveals

some critical distinctions. Even though it has fewer parameters and FLOPs, EfficientDet is significantly behind SPS-YOLOv5 in terms of FPS and Recall. A lower Recall suggests a higher probability of missing detections, and as shown in Figure 10, EfficientDet exhibits noticeable missed detections with smoke. In fire detection, missing detections can be extremely dangerous, rendering EfficientDet less effective for fire detection tasks compared to SPS-YOLOv5. The YOLOv3-EfficientNet [23] is based on YOLOv3 and is similar to EfficientDet. Although it has smaller parameters and FLOPs, its mAP<sup>50</sup> and Recall metrics are both subpar. In conclusion, SPS-YOLOv5 demonstrates the best overall performance among the compared algorithms and is well-suited for fire detection tasks, given its superior combination of mAP, FPS, and Recall.



(a)



(b)

**Figure 10.** Detection results of EfficientDet and SPS-YOLOv5: (**a**) Partial detection effect of EfficientDet; (**b**) Partial detection effect of SPS-YOLOv5.

## 5. Discussion

Forest fires can significantly impact the biological, chemical, and physical attributes of forest soil [48], thereby disrupting the ecological environment of the forest. This paper aims to provide an effective method for detecting wildfires. Nowadays, most wildfire detection methods can be categorized into those based on traditional machine learning and those based on computer vision [49]. The advantage of computer vision methods is their ability to learn multi-level representations of data. Compared to typical machine learning methods, it can better capture the complex structure of data and improve pattern recognition performance [50,51]. However, methods based on visual detection may require significant time for training due to large datasets and complex algorithms. Nevertheless, with the improvement in computer processing power, this issue has been mitigated to some extent. In summary, computer vision holds certain advantages in forest fire detection.

Unlike some stationary objects such as vehicles, safety helmets, masks, etc., fire is a dynamic entity. Furthermore, during a forest fire event, the texture, color, and dimensions of flames and smoke are highly complex, which makes the task of accurate fire detection extremely challenging. Therefore, it is necessary to improve the algorithm to make it suitable for the task of forest fire detection. In the comparative algorithms in this paper,

Faster R-CNN exhibits excellent performance in terms of Recall. It has a relatively large number of parameters and FLOPs, and a lower FPS, which is not conducive to real-time fire detection. CenterNet has fewer parameters than Faster R-CNN, and its mAP<sup>50</sup> and Recall performance are only average. SSD, as a classic one-stage detector, offers higher FPS and fewer parameters, but its mAP<sup>50</sup> and Recall are lower. These classic algorithms have shortcomings in both detection accuracy and speed. However, YOLOv5 effectively balances the trade-off between detection accuracy and speed. The experimental results show that YOLOv5 outperforms SSD in both mAP<sup>50</sup> and Recall, and it also boasts better FPS and fewer parameters than Faster R-CNN and CenterNet.

YOLOv5 tends to produce false positives when detecting images with complex smoke and flames. Therefore, in the backbone of YOLOv5, this paper introduces SAC to effectively fuse ordinary convolution and atrous convolution, enhancing the perception of the object's background environment, and thus enabling an accurate detection of smoke and fire. SAC, due to its use of atrous convolutions, significantly reduces the model's computational complexity and training time, as observed in the experimental results, which show an improvement in YOLOv5's mAP<sup>50</sup>. In a single image, the distribution of smoke and flames may be widely separated. To help the model capture a broader range of semantic information, this paper fuses convolution with PSA in the neck of YOLOv5. Experimental results indicate that the addition of PSA leads to a noticeable improvement in the model's mAP<sup>50</sup> and R. Given the widespread distribution of smoke and flames during a forest fire event, the issue of target overlap causing missed detections is alleviated by replacing the original NMS with Soft-NMS according to the detection results. When comparing to the previous most cutting-edge detection algorithms, such as EfficientDet and YOLOv3-EfficientNet, the latter two have fewer parameters and FLOPs, and their mAP<sup>50</sup> and Recall do not match the proposed SPS-YOLOv5. Furthermore, SPS-YOLOv5 also outperforms EfficientDet in terms of FPS.

Based on the results of the SPS-YOLOv5, it is suitable for fire detection compared to these algorithms. This study introduced a substantial amount of data for training, validation, and inference to achieve efficient and high-precision detection, with the main contributions being the following:

- Based on various drone images, we reconstructed a fire dataset that encompasses multiple scenarios and includes images of flames and smoke in different patterns.
- (2) By combining SAC, PSA, and Soft-NMS with YOLOv5, we designed and verified a more suitable fire detection algorithm: SPS-YOLOv5.
- (3) We compared SPS-YOLOv5 with several classical detection algorithms, and its good detection performance addresses the problem of complex targets detection and highlights the superiority of SPS-YOLOv5.

The experimental results indicate that SPS-YOLOv5 has broad applications in forest fire detection. In terms of practical deployment, it can be installed on fire monitoring towers or stations to monitor surrounding wildfires in real time. It can also be mounted on high-altitude devices such as drones to rapidly survey large areas and detect potential sources of fire. Additionally, fixed or mobile cameras can be set up in the forest, with SPS-YOLOv5 deployed within them. These deployment strategies need to be customized based on the specific forest terrain and requirements to enhance the sensitivity and accuracy of the fire detection. The proper placement of vision-based detection devices can cover a large area, a feature that contributes to cost control when applying SPS-YOLOv5. Deploying SPS-YOLOv5 into forest equipment allows for the continuous analysis of image data, promptly identifying potential signs of wildfires and improving monitoring efficiency. The algorithm can recognize smoke and flame characteristics through image analysis, achieving a high-precision detection of fire sources, reducing false alarms, and ensuring accuracy. This algorithm can automatically analyze image data, provide real-time notifications, and assist relevant organizations and personnel in rapidly responding to take necessary measures to extinguish the fire.

However, this paper's work also has certain limitations. In order to better deploy the algorithm, it is essential to effectively reduce the algorithm's parameters. This work tends to enhance the algorithm's detection precision and recall, inevitably increasing the algorithm's parameters. Moreover, in the forest, there are various interference factors. For example, in the forest, factors such as fog, red leaves, and sunlight during sunset can introduce interference. Due to the similarity in their color and texture information with smoke and flames, these factors can easily lead to false detections in the algorithm proposed in this paper. Therefore, in future work, we plan to collect images of fog, sunsets, and red leaves in the forest and create a more diverse dataset with richer backgrounds. This dataset will enable us to investigate the correlation between the shape characteristics of fires and their surrounding environments, thereby further enhancing our algorithm's capabilities for fire detection and resistance to interference. Furthermore, in the future, we are committed to achieving algorithm lightweighting while minimizing the sacrifice of algorithm performance.

#### 6. Conclusions

Since flames and smoke often exhibit irregular shapes, accurate and effective fire detection is still a challenging task. Detecting and identifying fire types and corresponding fire suppression measures in a timely manner can prevent fire spread and reduce economic losses. Therefore, it is important and meaningful to study the efficient detection of forest fires.

This paper proposes an improved YOLOv5 detection algorithm. First, we replace some convolutions in the YOLOv5 backbone with SAC to enhance the extraction of background information. Then, PSA is introduced to improve the algorithm's ability to extract feature information related to smoke and flames. Finally, Soft-NMS is employed to replace NMS for accurately detecting multiple overlapping targets. The experimental results show that the proposed algorithm achieves 79.8% mAP<sup>50</sup> and 43 FPS. Compared to the traditional YOLOv5, SPS-YOLOv5 achieves a better balance between precision, recall, and FLOPs, demonstrating superior detection performance. These improvements effectively improve the performance of SPS-YOLOv5. Due to the complex fire environment, multi-scale forest fire recognition is difficult. SPS-YOLOv5 has a good detection capability on multi-scale and multi-scene fire targets, especially the early and middle stage forest fire targets. When forest fire detection is performed, it can effectively protect the forest and people's lives and property.

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

**Funding:** This work is supported in part by the Jiangsu Province Natural Science Project of Institution (Grant No. 21KJB520020), National Natural Science Foundation of China (No. 42105143), 2021 the 14th Five-Year for Jiangsu Province Education Science Project (Grant No. B/2021/01/15), Wuxi Science and Technology Plan Project (Grant No. K20221044), Wuxi Science and Technology Plan Project (Grant No. K20231011), Wuxi "Xishan Talent Plan" Innovation leader Talent Project (Grant No. 2022xsyc002), and the Qing Lan Project of Jiangsu Province.

**Data Availability Statement:** The data used in this study can be obtained from the corresponding author upon reasonable request.

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

## References

- Guha-Sapir, D.; Hoyois, P.; Below, R. Annual Disaster Statistical Review 2015: The Numbers and Trends; Louvain-la-Neuve, Belgium, 2016; pp. 27–28. Available online: http://www.cred.be/sites/default/files/ADSR\_2015.pdf (accessed on 31 October 2016).
- Zhang, N.; Shen, S.L.; Zhou, A.N.; Chen, J. A brief report on the March 21, 2019 explosions at a chemical factory in Xiangshui, China. Process Saf. 2019, 38, e12060. [CrossRef]

- 3. Zhao, B. Facts and lessons related to the explosion accident in Tianjin Port. China Nat. Hazards 2016, 84, 707–713. [CrossRef]
- 4. Wu, L.; Chen, L.; Hao, X. Multi-Sensor Data Fusion Algorithm for Indoor Fire Early Warning Based on BP Neural Network. *Information* **2021**, *12*, 59. [CrossRef]
- 5. Hu, K.; Li, Y.; Xia, M.; Wu, J.; Lu, M.; Zhang, S.; Weng, L. Federated learning: A distributed shared machine learning method. *Complexity* **2021**, 2021, 8261663. [CrossRef]
- Solórzano, A.; Eichmann, J.; Fernández, L.; Ziems, B.; Jiménez-Soto, J.M.; Marco, S.; Fonollosa, J. Early fire detection based on gas sensor arrays: Multivariate calibration and validation. *Sens. Actuators B Chem.* 2022, 352, 130961. [CrossRef]
- Sun, B.; Xu, Z.D. A multi-neural network fusion algorithm for fire warning in tunnels. *Appl. Soft Comput.* 2022, 131, 109799. [CrossRef]
- 8. Pang, Y.; Li, Y.; Feng, Z.; Feng, Z.; Zhao, Z.; Chen, S.; Zhang, H. Forest fire occurrence prediction in China based on machine learning methods. *Remote Sens.* **2022**, *14*, 5546. [CrossRef]
- 9. Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]
- 10. Kalantar, B.; Ueda, N.; Idrees, M.O.; Janizadeh, S.; Ahmadi, K.; Shabani, F. Forest fire susceptibility prediction based on machine learning models with resampling algorithms on remote sensing data. *Remote Sens.* **2020**, *12*, 3682. [CrossRef]
- 11. Hu, K.; Wu, J.; Li, Y.; Lu, M.; Weng, L.; Xia, M. Fedgen: Federated learning-based graph convolutional networks for non-euclidean spatial data. *Mathematics* **2022**, *10*, 1000. [CrossRef]
- 12. Hu, K.; Weng, C.; Shen, C.; Wang, T.; Weng, L.; Xia, M. A multi-stage underwater image aesthetic enhancement algorithm based on a generative adversarial network. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106196. [CrossRef]
- Li, P.; Zhao, W. Image fire detection algorithms based on convolutional neural networks. *Case Stud. Therm. Eng.* 2020, 19, 100625. [CrossRef]
- 14. Zhao, E.; Liu, Y.; Zhang, J.; Tian, Y. Forest fire smoke recognition based on anchor box adaptive generation method. *Electronics* **2021**, *10*, 566. [CrossRef]
- 15. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; Volume 28.
- Nguyen, A.Q.; Nguyen, H.T.; Tran, V.C.; Pham, H.X.; Pestana, J. A Visual Real-time Fire Detection using Single Shot MultiBox Detector for UAV-based Fire Surveillance. In Proceedings of the 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE), Phu Quoc, Vietnam, 13–15 January 2021; Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 2021; pp. 338–343.
- Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A.C. SSD: Single Shot Multibox Detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; Part I 14; pp. 21–37.
- 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]
- 19. Zhang, Y.; Zhang, J.; Wang, Q.; Zhong, Z. DyNet: Dynamic Convolution for Accelerating Convolutional Neural Networks. *arXiv* **2020**, arXiv:2004.10694.
- Li, Y.; Zhang, W.; Liu, Y.; Jing, R.; Liu, C. An efficient fire and smoke detection algorithm based on an end-to-end structured network. *Eng. Appl. Artif. Intell.* 2022, 116, 105492. [CrossRef]
- Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. Comput. Vis. ECCV 2020, 2020, 213–229.
- Qin, Y.Y.; Cao, J.T.; Ji, X.F. Fire Detection Method Based on Depthwise Separable Convolution and YOLOv3. Int. J. Autom. Comput. 2021, 18, 300–310. [CrossRef]
- 23. Redmon, J.; Farhadi, A. Yolov3: An Incremental Improvement. arXiv 2018, arXiv:1804.02767.
- 24. Wang, Y.; Hua, C.; Ding, W.; Wu, R. Real-time detection of flame and smoke using an improved YOLOv4 network. *Signal Image Video Process.* **2022**, *16*, 1109–1116. [CrossRef]
- Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
- 26. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* 2020, arXiv:2004.10934.
- 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, 14–19 June 2020; pp. 10781–10790.
- 28. Bahhar, C.; Ksibi, A.; Ayadi, M.; Jamjoom, M.M.; Ullah, Z.; Soufiene, B.O.; Sakli, H. Wildfire and Smoke Detection Using Staged YOLO Model and Ensemble CNN. *Electronics* **2023**, *12*, 228. [CrossRef]
- Yu, S.; Sun, C.; Wang, X.; Li, B. Forest fire detection algorithm based on Improved YOLOv5. J. Phys. Conf. Ser. 2022, 2384, 012046. [CrossRef]
- Dou, Z.; Zhou, H.; Liu, Z.; Hu, Y.; Wang, P.; Zhang, J.; Wang, Q.; Chen, L.; Diao, X.; Li, J. An Improved YOLOv5s Fire Detection Model. *Fire Technol.* 2023. [CrossRef]
- 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.

- 32. Du, H.; Zhu, W.; Peng, K.; Li, W. Improved High Speed Flame Detection Method Based on YOLOv7. *Open J. Appl. Sci.* 2022, 12, 2004–2018. [CrossRef]
- Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.
- Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv 2022, arXiv:2207.02696.
- 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]
- Li, H.; Li, J.; Wei, H.; Liu, Z.; Zhan, Z.; Ren, Q. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. arXiv 2022, arXiv:2206.02424.
- 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.
- 38. Talaat, F.M.; ZainEldin, H. An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Comput. Appl.* **2023**, 35, 20939–20954. [CrossRef]
- Qiao, S.; Chen, L.C.; Yuille, A. DetectoRS: Detecting Objects with Recursive Feature Pyramid and Switchable Atrous Convolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10213–10224.
- 40. Liu, H.; Liu, F.; Fan, X.; Huang, D. Polarized Self-Attention: Towards High-quality Pixel-wise Regression. *arXiv* 2021, arXiv:2107.00782.
- Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS—Improving Object Detection with One Line of Code. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5561–5569.
- 42. Neubeck, A.; Van Gool, L. Efficient Non-Maximum Suppression. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; IEEE: Piscataway, NJ, USA, 2006; Volume 3, pp. 850–855.
- Wang, C.; Liao, H.M.; Wu, Y.; Chen, P.; Hsieh, J.; Yeh, I. 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; Springer: Cham, Switzerland; pp. 390–391.
- 44. Lin, T.Y.; Dollar, 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.
- 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.
- Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmen-tation with deep convolutional nets and fully connected crfs. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
- 47. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as Points. arXiv 2019, arXiv:1904.07850.
- 48. Agbeshie, A.A.; Abugre, S.; Atta-Darkwa, T.; Awuah, R. A review of the effects of forest fire on soil properties. J. For. Res. 2022, 33, 1419–1441. [CrossRef]
- 49. Chew, Y.J.; Ooi, S.Y.; Pang, Y.H.; Wong, K.S. A Review of forest fire combating efforts, challenges and future directions in Peninsular Malaysia, Sabah, and Sarawak. *Forests* **2022**, *13*, 1405. [CrossRef]
- 50. Alkhatib, R.; Sahwan, W.; Alkhatieb, A.; Schütt, B. A Brief Review of Machine Learning Algorithms in Forest Fires Science. *Appl. Sci.* **2023**, *13*, 8275. [CrossRef]
- 51. 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]

**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.