

Article

Forest Fire Smoke Recognition Based on Anchor Box Adaptive Generation Method

Enting Zhao ^{1,2,†} , Yang Liu ^{1,2,†} , Junguo Zhang ^{1,2,*} and Ye Tian ^{1,2,*}

¹ School of Technology, Beijing Forestry University, Beijing 100083, China; zhao_enting@163.com (E.Z.); liuyang_yyy@126.com (Y.L.)

² Key Lab of State Forestry and Grassland Administration for Forestry Equipment and Automation, Beijing 100083, China

* Correspondence: zhangjunguo@bjfu.edu.cn (J.Z.); tytoemail@bjfu.edu.cn (Y.T.)

† These authors contributed equally to this work and should be considered co-first authors.

Abstract: There are major problems in the field of image-based forest fire smoke detection, including the low recognition rate caused by the changeable and complex state of smoke in the forest environment and the high false alarm rate caused by various interferential objects in the recognition process. Here, a forest fire smoke identification method based on the integration of environmental information is proposed. The model uses (1) the Faster R-CNN as the basic framework, (2) a component perception module to generate a receptive field of integrated environmental information through separable convolution to improve recognition accuracy, and (3) a multi-level Region of Interest (ROI) pooling structure to reduce the deviation caused by rounding in the ROI pooling process. The results showed that the model achieved a recognition accuracy rate of 96.72%, an Intersection Over Union (IOU) of 78.96%, and an average recognition speed for each picture of 1.5 ms; the false alarm rate was 2.35% and the false-negative rate was 3.28%. Compared with other models, the proposed model can effectively enhance the recognition accuracy and recognition speed of forest fire smoke, which provides a technical basis for the real-time and accurate detection of forest fires.

Keywords: forest fire smoke; Faster R-CNN; receptive field; anchor box; multi-level ROI pooling



check for updates

Citation: Zhao, E.; Liu, Y.; Zhang, J.; Tian, Y. Forest Fire Smoke Recognition Based on Anchor Box Adaptive Generation Method.

Electronics **2021**, *10*, 566.

<https://doi.org/10.3390/electronics10050566>

electronics10050566

Academic Editor: Dah-Jye Lee

Received: 31 January 2021

Accepted: 24 February 2021

Published: 27 February 2021

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



Copyright: © 2021 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

Forest fire is one of the major environmental disasters threatening the safety of forestry workers and the ecological balance. From January to August 2019, a total of 1563 forest fires occurred in China, which affected an area of 8518 hectares and resulted in substantial ecological damage and economic losses. The effective prevention and control of forest fires have thus become a major focus of scientific research. The early detection and warning of fires is critically important for fire prevention. When a forest fire occurs, it is often accompanied by smoke. Given that smoke can be more easily detected because of its wider distribution compared to flames, it is a robust indicator for the presence of forest fires in the early stages. By detecting smoke, forest fires can be detected so that fire-fighting measures can be quickly implemented, and the harm caused by forest fires can be reduced.

Forest fire smoke recognition based on computer vision is the most widely used method for detecting forest fires. Zheng Huaibing et al. [1] combined the characteristics of smoke to design a variety of static and dynamic features and detect moving targets through the background difference method. In normal weather and on foggy days the recognition accuracy rates reached 92.7% and 76.3%, respectively. Surt et al. [2] proposed a digital image processing method that detects fire smoke by using the convex hull algorithm to calculate the area of interest and segments the changing area from the image. Zhao Min et al. [3] developed the smoke detection algorithm based on multiple texture features and used an SVM (Support Vector Machine) classifier for classification, which effectively improved the detection rate of video smoke detection methods. Li Hongdi et al. [4] developed a smoke

detection algorithm, using image pyramid texture and edge multi-scale features, which could achieve a detection rate of more than 94%; the false alarm rate is less than 3.0% on an image set with more than 80% smoke. Although the above fire smoke recognition methods based on traditional technologies realize high recognition rates, they cannot effectively use background information to identify smoke in forests with complex backgrounds and greater interference.

2. Related Work

With the fast development of artificial intelligence technology, image recognition technology based on machine learning has been broadly applied to smoke recognition. Wang Tao et al. [5] used a method that combined traditional manual feature extraction and CNN's automatic extraction of smoke image features. The recognition accuracy of the method reached 92% and the false alarm rate was 3.3%. Yin et al. [6] developed a deep normalized convolutional neural network for fire smoke detection, which used a batch of normalization methods to accelerate the training process and enhance the accuracy of smoke detection. Various target detection models such as Denet [7], Light-Head [8], Yolo-V3 [9], and Faster R-CNN [10] have been applied to forest fire smoke detection. Zhang et al. [11] used Faster R-CNN to detect wild forest fire smoke by inserting both real smoke and simulated smoke into the forest to generate synthetic smoke images, which could not only avoid the complicated artificial feature extraction process but also ease the training data and avoid the problem of scarcity. Lin et al. [12] developed an improved smoke recognition model with non-maximum annexation based on Faster R-CNN and 3D CNN, which achieves fire smoke recognition by combining dynamic spatiotemporal information. The recognition accuracy of the smoke video was 95.23%, and the false alarm rate was 0.39%. Li et al. [13] applied Candidate Smoke Region Segmentation and the proposed wildfire smoke dilated DenseNet (WSDD-Net) to forest fire smoke detection, which realized a high AR of 99.20% and a low false-negative rate (MA) of 0.24%. The above fire smoke recognition model achieved higher recognition accuracy and a lower false-negative rate than traditional machine vision algorithms. With the rapid development of science and technology, the field of target detection has become increasingly applied to species recognition. However, slow recognition speed and low recognition efficiency are still bottlenecks that need to be further resolved. With the continuous development of the forest fire smoke recognition algorithm, it is found that the background of the image has a certain impact on the accuracy of algorithm, but most of the current algorithms have not incorporated more background to improve it.

Current smoke detection methods based on deep convolutional networks mainly consist of one-stage and two-stage methods. The former directly outputs the location of the smoke while the latter is divided into two steps. First, the coarse positions of all possible locations of the smoke are located. After the possible locations are determined, the corrected results for high-probability locations are generated. The one-stage method is faster than the two-stage method, but there is a risk of omission. Moreover, for smoke detection, the results of missed detection can lead to serious consequences. Therefore, the two-stage method is more effective for forest fire smoke detection. Faster R-CNN is the typical model of two-stage approach. Based on previous research, this article introduces multi-scale receptive fields and multi-level ROI pooling in target detection networks that can incorporate more environmental background information in the identification process. This can reduce the influence of interference factors such as clouds and fog to improve the extraction ability of fire smoke features and build a new fire smoke recognition method to achieve high efficiency and high accuracy for fire smoke recognition.

3. Materials and Methods

3.1. Construction of the Forest Fire Smoke Image Set

A total of 24 fire smoke videos and 9899 high-definition images were obtained from fire monitoring manufacturers, other research groups, and related governmental departments.

The smoke images varied and ranged from far-sighted fire smoke to near-sighted fire smoke. The selected smoke samples included 5502 high-definition images containing fire smoke, and 2150 images including clouds, fog, and haze were also collected as non-smoke samples, which together constituted our training dataset. Typical smoke images and interference images are shown in Figure 1.

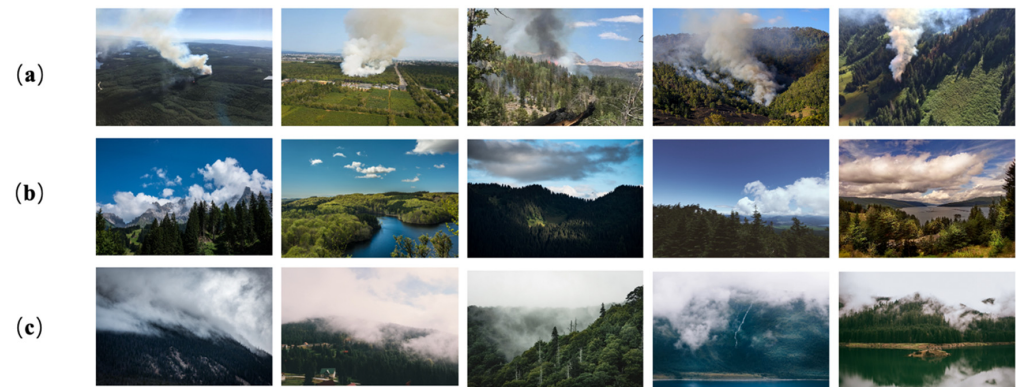


Figure 1. Smoke and interference images. (a) Smoke images, (b) cloud images, (c) fog images.

3.2. Faster R-CNN

The optimized Faster R-CNN network is shown in Figure 2. We used ResNet50 as the backbone as ResNet50 has more convolutional layers and residual structure, which can merge more feature information and learn more effective feature information with higher dimensions. The model uses the Region Proposal Network (RPN) [14] to generate candidate regions that are input into the pooling layer with the feature layer to achieve convolutional layer feature sharing. The actual candidate area was obtained by regression of the anchor box, and the anchor box of fixed size corresponded to the receptive field of the corresponding size. To ensure the universality of the target detection results, the anchor box was typically set to 128×128 , 256×256 , and 512×512 to adapt to targets of different sizes; in addition, scale ratios of 1:1, 1:2, and 2:1 were used to adapt to objects with different shapes. After anchor boxes had pre-defined scales and aspect ratios, the first step of the training sample generation process was to filter out the anchor boxes that exceeded the image boundaries. For the remaining anchors, positive samples were those with Intersection Over Union (IOU) overlap ratios with ground-truth bounding boxes higher than 0.7. Next, 256 anchors were randomly selected from the positive and negative samples to form a minibatch for training.

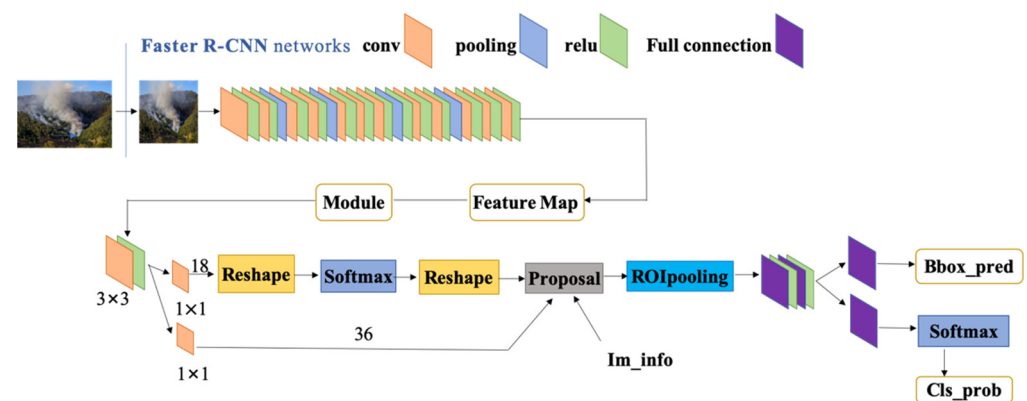


Figure 2. Structure of Faster R-CNN.

3.3. Integration of Environmental Information into Receptive Fields

In this study, a method was proposed in which the component perception modules are used to integrate environmental information into receptive fields. Specifically, the

component perception module was used to extract the characteristic information of each area block. This module can expand the receptive field and therefore capture more environmental information. Figure 3 shows the RPN structure after inserting the sensing component, where Awareness Module (AM) is the component sensing module as shown in Figure 4. The AM module used an attention mechanism to focus on the characteristics of the object while considering environmental information. This study first used a 1×1 convolution kernel to reduce the dimensionality. To improve the speed, a separable convolution method was used, and the convolution kernels were $h_i \times 1$ and $1 \times w_i$ respectively for convolution operations. Finally, the convolution operation with the convolution kernel of 1×1 was used again to increase the feature dimension to the original dimension, and the final result was outputted by element-wise summation with the original input. Therefore, the overall structure was residual. This enabled the environmental information to be integrated into the expression of regional features to further enhance the network's ability to extract target features.

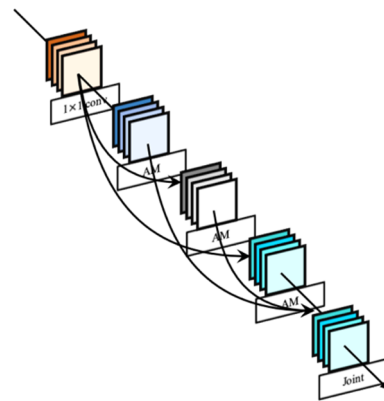


Figure 3. Overall structure diagram.

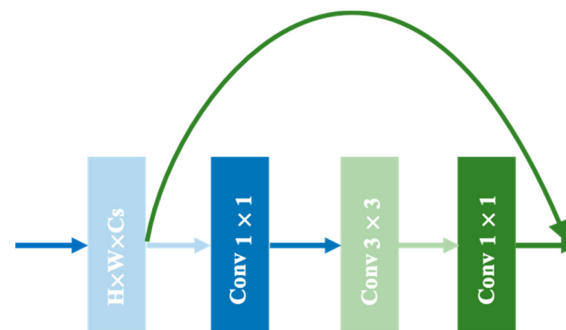


Figure 4. Component awareness module.

3.4. Multi-Level ROI Pooling Layer

We used the idea of the spatial pyramid pooling structure [15]. In Faster R-CNN algorithm, the multi-level ROI Pooling layer optimization is applied after the convolutional layer and before classification and regression, and its function is to generate a recommended feature map with fixed size and obtain fixed output data as the input of the fully connected layer. As there are two rounds in the ROI pooling layer, pixels are lost in some areas, which causes deviation in the position of the suggestion frame and affects the accuracy of the fire smoke detection results. A spatial pooling pyramid was introduced to divide the candidate frame into regions (such as 1×1 , 2×2 , and 3×3) in different spaces and then connect them into a single vector for output. Figure 5 is a schematic diagram of spatial pyramid pooling. Spatial pyramid pooling can take both coarse-scale and fine-scale information into account simultaneously, thus avoiding the deviation caused by the double rounding of the single-level ROI pooling layer and improving the accuracy of fire smoke

detection. At the same time, the spatial pooling pyramid is more robust to deformable and variable-scale features.

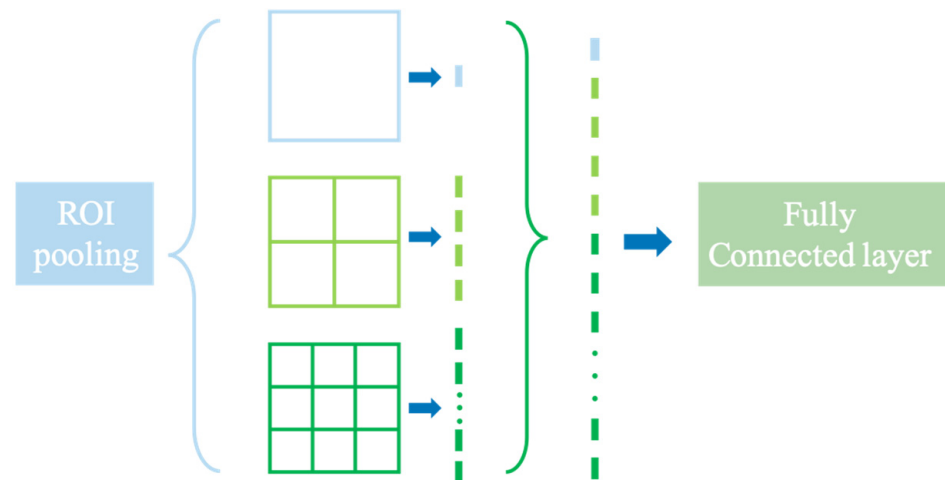


Figure 5. Multi-level ROI pooling layer.

Although the multi-level ROI pooling layer improves the accuracy of fire and smoke detection, it also increases the network parameters and the network training time. Therefore, a suitable division scheme needs to be used for ROI pooling at all levels. Table 1 lists three different three-level ROI pooling layer division schemes and the number of grid parameters selected in this article. Table 1 is a single-level ROI pooling division scheme. It can be seen from Table 1 that the precision of scheme 3 is 96.72%, which is superior to schemes 1 and 2, so scheme 3 was selected to meet the requirements of detection speed and recognition accuracy simultaneously.

Table 1. ROI pooling layer division scheme.

Methods	Classification Method	Mesh Parameter	Precision	Speed (ms)
1	Original 7×7	$7 \times 7 \times 512 \times 4096$	96.07%	0.018
2	$1 \times 1+2 \times 2+4 \times 4$	$21 \times 512 \times 4096$	96.37%	0.020
3	$2 \times 2+4 \times 4+8 \times 8$	$84 \times 512 \times 4096$	96.72%	0.026

4. Results

4.1. Experimental Platform

Tensorflow1.0 was used as the deep learning framework in this experiment, and dual GPUs were used to accelerate the training process. The hardware configuration consisted of an Intel Core i5-3210M processor and NVIDIA GeForce GTX 1080Ti graphics card; the software environment included ubuntu 16.04LTS and python3.6.0. The GPU acceleration library used CUDA8.0 and CUDNN5.1.

4.2. Image Annotation

The types and numbers of image used in the study are shown in Table 2.

Table 2. Forest fire smoke dataset.

Dataset	Smoke Image	Interference Image	Total	Application
PART 1	3332	1290	4622	training
PART 2	1110	430	1540	validation
PART 3	1110	430	1540	test

The identification of fire smoke was realized by target detection, and the label image tool was used to make the Visual Object Classes (VOC) dataset file.

4.3. Evaluation Index

To better evaluate the accuracy of forest fire smoke recognition, four indicators were used herein, including pixel accuracy, category average accuracy, average intersection ratio, and frequency-weighted average intersection ratio for model performance evaluation. Larger values of the four indicators corresponded to superior recognition effects.

Precision (P) and Recall (R) are the two simplest evaluation indicators, which represent the proportion of correctly classified images out of the total number of images and the number of correctly classified images out of the images that should be correctly classified. The specific equations are shown below.

$$P = \frac{TP}{TP + FN} \quad (1)$$

$$R = \frac{TP}{TP + FP} \quad (2)$$

The Mean Average Precision (MAP) provides a comprehensive measure of the average accuracy of the detected target, and it indicates the average of each category of Average Precision (that is, the average accuracy of all categories is summed and divided by all categories). The specific equation for MAP is as follows:

$$MAP = \frac{\sum \text{AveragePrecision}}{N(\text{Classes})} \quad (3)$$

Intersection Over Union (IOU) is the most commonly used evaluation index in the object detection field and is used as a standard for measuring the accuracy of detecting corresponding objects in a specific dataset. IOU represents the overlap rate between the candidate box and the ground-truth bound, namely the ratio of their intersection and union. Higher intersection ratios correspond to higher correlations. Figure 6 illustrates the cross-to-bin ratio.

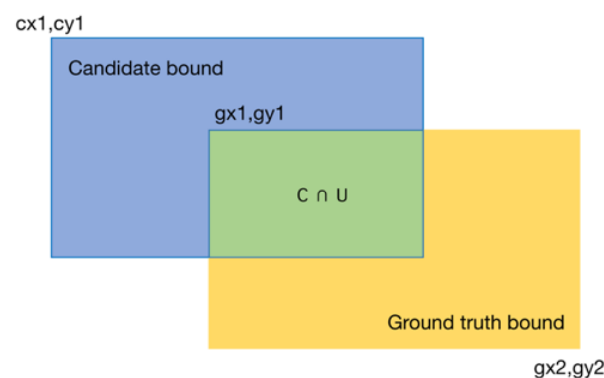


Figure 6. Intersection over union.

The equation for the cross-to-bin ratio is as follows:

$$Iou = \frac{\text{area}(C) \cap \text{area}(G)}{\text{area}(C) \cup \text{area}(G)} \quad (4)$$

In addition, two indicators were specifically used to evaluate fire detection performance, namely the false-negative rate and the false alarm rate.

The false-negative rate (MA) is the number of unrecognized images divided by total number of images.

$$MA = \frac{FN}{TP + FN} \quad (5)$$

The false alarm rate (FA) is the number of incorrectly recognized images divided by the total number of images.

$$FA = \frac{FP}{FP + TN} \quad (6)$$

In the equations, TP and TN are the number of true positives and true negatives, respectively, and FP and FN are the number of false positives and false negatives, respectively.

4.4. Forest Fire Smoke Detection Results

The results of ablation experiments are as described in Table 3. The recognition accuracy of Faster R-CNN is 94.48%, the recognition accuracy of Faster R-CNN + Component awareness module is 96.07%, and Faster R-CNN + Multi-level ROI Pooling is 95.23%, and the recognition accuracy of Faster R-CNN + Component awareness module + Multi-level ROI Pooling is 96.72%.

Table 3. Ablation Study.

Model	Faster R-CNN	Faster R-CNN + Component Awareness Module	Faster R-CNN + Multi-level ROI	Faster R-CNN + Component Awareness Module + Multi-level ROI
Precision	94.48%	96.07%	95.23%	96.72%

After comparison, it can be seen that after adding the Component awareness module, the recognition accuracy rate increased by 1.59% compared with Faster R-CNN, which proves the effectiveness of the Component awareness module.

After adding Multi-level ROI Pooling, the recognition accuracy rate increased by 0.75% compared with Faster R-CNN, which proves the effectiveness of Multi-level ROI Pooling.

After adding the Component awareness module and Multi-level ROI Pooling, the model had an accuracy of 96.72%, which proves that the module strengthens important features and attenuates unimportant features; multi-level ROI can detect and recognize targets more effectively.

The results of the method used in this article to identify forest fire smoke are shown in Figure 7. Figure 7 shows that the recognition accuracy of this method can be greater than 96% for both far-sighted fire smoke and near-sighted fire smoke.

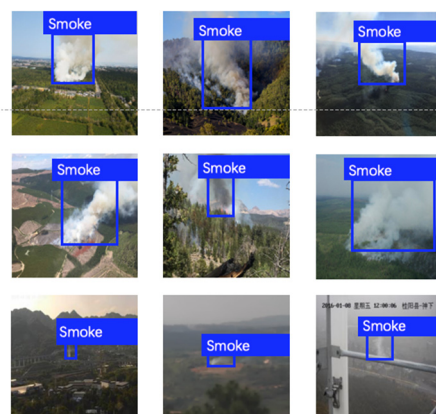


Figure 7. Recognition effect of the proposed model.

5. Discussion

5.1. Comparative Experiment of Different Recognition Models

To verify the superiority of the forest fire smoke recognition model proposed in this article and compare it with the recognition effects of two commonly used target detection

models (Yolo-V3 and SSD) and the latest target detection models (FCOS and EfficientDet) with Faster R-CNN, 1110 images of fire smoke at different viewing distances in the dataset PART 3 were selected. The focal image and 430 interfering images, such as clouds and fog, were used as identification objects. Figure 8 shows the comparison of the recognition effects of these three models. Several of the relevant findings are detailed below.

1. The model proposed in this article achieved a higher recognition accuracy and MAP, while the Yolo-V3 model had the lowest recognition accuracy and MAP. The recognition accuracy difference between the two was 17.36%, and the MAP difference was 6.81%.
2. The model proposed in this article realized good performance in terms of cross-combination, but the differences among the five models were not large. This finding likely stems from the fact that smoke has various shapes and the accurate marking of the smoke area is difficult. Consequently, the intersection ratio is generally not high.
3. The Yolo-V3 model had the lowest performance in terms of the false-negative rate and false alarm rate. The false alarm rate exceeded 10% and the false-negative rate exceeded 20%. For the proposed model, the false-negative and false alarm rates were both less than 4%, and the false-negative rate was lower (3.28%). Thus, the performance of fire smoke detection was higher.
4. We chose FCOS and EfficientDet methods for comparison, which are representative. The recognition accuracy of FCOS and EfficientDet is 97.03% and 97.65%, respectively. Although these two latest algorithms perform a little better than our recognition in accuracy (except for smoke image recognition with unclear smoke contour features and a small proportion of smoke images) FCOS and EfficientDet have the phenomenon of underreporting, while our algorithm can effectively detect the smoke in images.

As mentioned previously, Faster R-CNN based on a two-stage algorithm not only has a higher recognition accuracy than Yolo-V3, but also achieves comparable performance with the recent FCOS and EfficientDet methods. In addition, the improvement proposed in this article can better extract the features of smoke, which performs better than the unimproved Faster R-CNN and achieves a robust recognition effect.

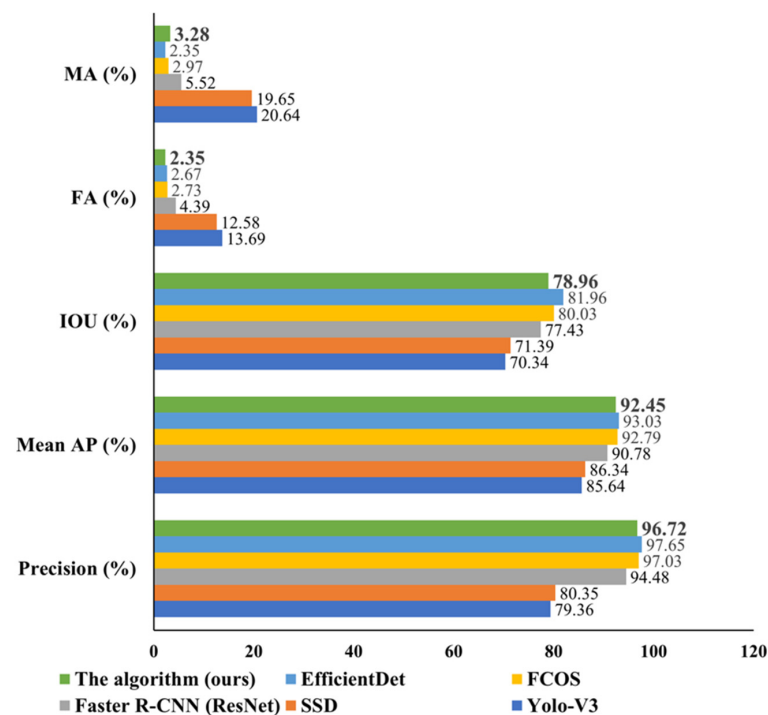


Figure 8. Detection accuracy of each algorithm.

5.2. Evaluation of Recognition Speed

The speed of fire smoke detection is one of the most important indicators for evaluating a recognition model. The samples in the verification set were used for testing, and the average recognition time of each image was calculated (Table 4). The recognition speed of the model proposed in this article was 0.041 ms faster than the Faster R-CNN algorithm before optimization. The average speed of smoke recognition was 0.003 ms, which is slower than that of the Yolo-V3 algorithm; nevertheless, in the actual process of forest fire smoke recognition, the recognition speed of the model proposed in this article has already met the target detection requirement.

Table 4. Single image running time for each algorithm.

Model	Speed (ms)
Yolo-V3	0.023
SSD	0.021
FCOS	0.019
EfficientDet	0.016
Faster R-CNN(ResNet)	0.067
Our algorithm	0.026

6. Conclusions

In this study, the anchor box generation mechanism in the model RPN network was optimized by circumventing the limitations associated with the fixed anchor box receptive field size and the inability to accurately sense the target. The perception component module was set to automatically generate an anchor box with a suitable receptive field size. At the same time, a multi-level ROI pooling layer was used to enhance the accuracy of fire smoke detection. For actual fire smoke images, the recognition accuracy rate was over 96%, the false-negative rate was reduced to 3.28%, and the detection speed reached 0.026 ms. The optimized smoke recognition method did not require a specific candidate box search algorithm. The entire network model can complete end-to-end detection tasks. Multi-level pooling can enhance the feature extraction ability of the network and improve the accuracy of forest fire smoke recognition. Reducing the rate of underreporting improves the informatization level of forest fire monitoring and increases its utility.

Author Contributions: Data curation, Y.L.; project administration, J.Z.; software, E.Z.; supervision, Y.T.; writing—original draft, E.Z. and Y.L.; writing—review and editing, E.Z. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by the Fundamental Research Funds for the Central Universities (Grant No. 2016ZCQ08 and Grant No. 2019SHFWLC01).

Institutional Review Board Statement: The authors declare Ethical review and approval were waived for this study, due to this manuscript is just general scientific research and does not involve moral issues.

Informed Consent Statement: The authors declare Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the research group's forest fire smoke identification related research is still being carried on, the later work will also rely on the current data set to carry out.

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

References

1. Zheng, H.B.; Zhai, J.Y. Forest fire smoke detection based on video analysis. *J. Nanjing Univ. Sci. Technol.* **2015**, *39*, 686–691.
2. Surit, S.; Chatwiriya, W. Forest Fire Smoke Detection in Video based on Digital Image Processing Approach with Static and Dynamic Characteristic Analysis. In Proceedings of the IEEE Conference on Computers 2011, Barcelona, Spain, 6–13 November 2011; pp. 35–39.
3. Zhao, M.; Zhang, W.; Wang, X.; Liu, Y. A Smoke Detection Algorithm with Multi-Texture Feature Exploration Under a Spatio-Temporal Background Model. *J. Xi'an Jiaotong Univ.* **2018**, *52*, 67–73. Available online: https://xueshu.baidu.com/usercenter/paper/show?paperid=1u4r0tb0ge4e08f0173k0a006k437807&site=xueshu_se&hitarticle=1&sc_from=nenu (accessed on 24 February 2021).
4. Li, H.D.; Yuan, F.N. Image based smoke detection using pyramid. texture and edge features. *J. Image Graph.* **2015**, *20*, 772–780.
5. Wang, T.; Gong, N.S.; Jiang, G.X. Smoke recognition based on the depth learning. *Comput. Technol. Appl.* **2018**, *44*, 137–141.
6. Yin, Z.J.; Wan, B.Y.; Yuan, F.N.; Xia, X.; Shi, J.T. A Deep Normalization and Convolutional Neural Network for Image Smoke Detection. *IEEE Access* **2017**, *5*, 18429–18438. [[CrossRef](#)]
7. Tychsen-Smith, L.; Petersson, L. DeNet: Scalable Real-time Object Detection with Directed Sparse Sampling. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; Volume 54, pp. 428–436.
8. Li, Z.; Peng, C.; Yu, G.; Zhang, X.Y.; Deng, Y.D.; Sun, J. Light-Head R-CNN: In Defense of Two-Stage Object Detector. *arXiv* **2017**, arXiv:1711.07264.
9. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
10. Ren, S.Q.; He, K.M.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
11. Zhang, Q.X.; Lin, G.H.; Zhang, Y.M.; Xu, G.; Wang, J.J. Wildland forest fire smoke detection based on Faster R-CNN using synthetic smoke images. *Procedia Eng.* **2018**, *211*, 441–446. [[CrossRef](#)]
12. Lin, G.H.; Zhang, Y.M.; Xu, G.; Zhang, Q.X. Smoke detection on video sequences using 3 D convolutional neural networks. *Fire Technol.* **2019**, *55*, 1827–1847. [[CrossRef](#)]
13. Li, T.T.; Zhao, E.T.; Zhang, J.G.; Hu, C.H. Detection of Wildfire Smoke Images based on a Densely Dilated Convolutional Network. *Electronics* **2019**, *8*, 1131. [[CrossRef](#)]
14. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
15. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)] [[PubMed](#)]