Article

ForestFireDetector: Expanding Channel Depth for Fine-Grained Feature Learning in Forest Fire Smoke Detection

Long Sun, Yidan Li and Tongxin Hu *

Key Laboratory of Sustainable Forest Ecosystem Management-Ministry of Education, College of Forestry, Northeast Forestry University, 26 Hexing Road, Harbin 150040, China; sunl-cf@nefu.edu.cn (L.S.); 317000190@nefu.edu.cn (Y.L.)
* Correspondence: htxhtx@nefu.edu.cn

Abstract: Wildfire is a pressing global issue that transcends geographic boundaries. Many areas, including China, are trying to cope with the threat of wildfires and manage limited forest resources. Effective forest fire detection is crucial, given its significant implications for ecological balance, social well-being and economic stability. In light of the problems of noise misclassification and manual design of the components in the current forest fire detection model, particularly the limited capability to identify subtle and unnoticeable smoke within intricate forest environments, this paper proposes an improved smoke detection model for forest fires utilizing YOLOv8 as its foundation. We expand the channel depth for fine-grain feature learning and retain more feature information. At the same time, lightweight convolution reduces the parameters of the model. This model enhances detection accuracy for smoke targets of varying scales and surpasses the accuracy of mainstream models. The outcomes of experiments demonstrate that the improved model exhibits superior performance, and the mean average precision is improved by 3.3%. This model significantly enhances the detection ability while also optimizing the neural network to make it more lightweight. These advancements position the model as a promising solution for early-stage forest fire smoke detection.

Keywords: forest fire smoke detection; YOLOv8; expanding channel depth; SPD-Conv; GSConv

1. Introduction

The forest is the largest terrestrial ecosystem on Earth and serves as a crucial material foundation for the survival and progress of humanity [1]. China is a country with limited forest resources, with a forest coverage of 24.02%, and is currently undergoing a transitional stage where the ecological, social and economic benefits of forests are being harmonized with sustainable development in an all-around way. Diseases and insect pests caused by natural factors and frequent forest fires have a significant impact on forest resources. Among these factors, forest fires are dominant [2]. The majority of forest fires are initiated via natural phenomena like lightning, high temperatures and strong winds. However, some are also caused by human factors such as production fires, sacrificial fires and outdoor smoking. Once an uncontrollable forest fire breaks out, it disrupts the ecological balance, destroys infrastructure and poses a threat to life safety. Consequently, it has become a major forestry disaster worldwide. In recent years, owing to the combined effects of global climate change and human interventions, the frequency and duration of forest fires have increased [3]. The increase in forest fires has placed a greater strain on fire prevention and control efforts in China. Despite an increase in investment towards forest fire prevention, there has been no significant change in the area affected by fires [4]. In response, China has adopted a forest fire prevention policy that prioritizes preventive measures and active elimination as necessary. In addition to preventing fires at their source, it is of paramount importance to extinguish the fire rapidly in its early stages. Forest fire detection is a pivotal means that can enable the early identification of forest fires, prevent their spreading and thereby minimize the resulting losses.

Conventional forest fire detection methods primarily hinge on the deployment of sensor and infrared detection equipment within forested regions, offering a high degree of accuracy [5,6]. Nonetheless, these detectors exhibit certain limitations, such as delayed equipment responses caused by smoke diffusion and temperature increases. Notably, substances like methanol and ethanol, when combusted, release minimal smoke that often falls below the concentration threshold for detection. Forest fire detection requires accurate judgment as well as a rapid response in order to take timely action to combat the fire. However, traditional methods often have poor real-time performance due to various limitations. Researchers have conducted extensive research into image recognition with the rapid evolution of electronics and image technology [7]. Compared to traditional image detection methods, deep-learning models automatically extract features from the original image and classify and recognize them during the training process [8]. This approach offers high accuracy and real-time performance, giving it a significant advantage. While current efforts to detect forest fires using image-based methods primarily concentrate on identifying flames [9–11], their effectiveness is hampered by challenges such as light obstructions, which can sometimes render them unrecognizable to monitoring systems. During the initial stages of a fire, characteristics like smoldering and upward-drifting smoke dynamics lead to constant changes in its shape [12]. This volatility makes it both feasible and imperative to detect smoke early.

At present, both domestically and internationally, the detection of early forest fire is generally a one-stage model with regression, represented via YOLO [13–15] and SSD [16]. The researchers compared different models and found that YOLO had poor performance in detecting small targets [17–20]. Although SSD had better performance, it was not as good as Faster R-CNN [21]. To address these limitations, various attempts have been made to refine existing models. For instance, Zhao et al. [22] proposed an improved version of the Fire-YOLO model that concentrates on enhancing feature propagation for small targets and rectifying the subpar prediction capabilities of YOLO for small smoke instances. However, YOLOv3 uses a large receptive field to detect the target. There may be some instances where background noise is mistakenly identified as smoke, resulting in an increase in false alarms caused by noise interference. Rahman et al. [23] employed the SSD model, utilizing texture and color, to the UAV with high speed and accuracy. However, this approach remains susceptible to misjudgments involving clouds and smoke, and the manual setup of priority box dimensions and shapes necessitates debugging reliant on experience.

All the models mentioned above require a predefined set of anchors to adequately cover targets with different scales and aspect ratios. However, this fixed setting cannot fully adapt to the changes in different scales and shapes. Additionally, forest fire detection scenarios exhibit a disparity between negative and positive samples, potentially leading to model performance degradation due to sample imbalance. In this paper, we opted for YOLOv8, which has an anchor-free Head and a task-aligned assigner for loss calculation, as the foundational model. To enhance the smoke detection ability of the model, we expand the channel depth, which reduces information loss and improves feature extraction. In order to maintain accuracy while reducing the model parameters, we use a lightweight convolution instead of a standard convolution. Furthermore, we increase an additional detection head for small targets to maximize the utilization of the shallow information to improve the detection accuracy. Our paper makes the following contributions:

- We propose an improved model that expands the channel depth to reduce fine-grain loss and retains more feature information, thus enhancing the detection accuracy and providing a new idea and method for forest fire detection.
- This model makes the network more focused on the identification of small targets, improves the detection ability of small smoke and makes it more suitable for early-stage forest fires.
- The parameters of the model have been further reduced to 2.7M compared to the 3M of YOLOv8. It strikes a balance between computational efficiency and performance enhancement, making it possible for deployment.

The subsequent sections of this paper are organized in the following manner. Section 2 introduces the model structure, new modules and improved model in detail. In Section 3, we show the configuration of the experiment, the setting of some main training parameters, the experimental results of each improved part and a comparative analysis of the module gain in forest fire smoke detection. We describe the discussion and analysis of the model and the outlook for future work in Section 4. The complete research is summarized in Section 5.

2. Materials and Methods

2.1. Dataset

The quality of the datasets and the dimensions of the target influence the experimental results. In this paper, forest fire smoke images were taken using pan-tilt device and captured images of different locations, times and environmental circumstances. These images were labeled with Labelme and subsequently translated into the COCO format. Converting the data from json to the required txt format for YOLO compatibility followed this step. To enhance the dataset's quality, these images are preprocessed via data augmentation, achieving a twofold purpose of feature enhancement and data expansion. The dataset contains 3966 images: 3173 are training sets, and the rest are validation sets and test sets. Typical images are shown in Figure 1.
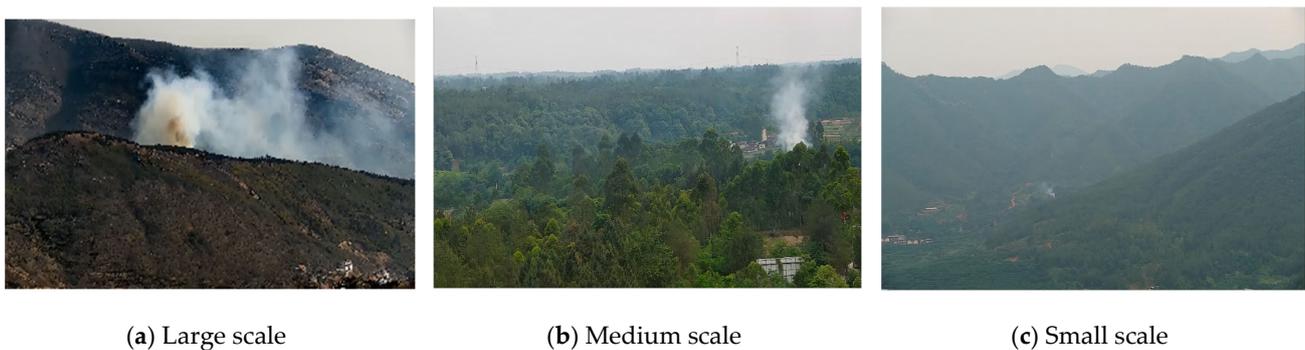

(**a**) Large scale　　　　　　(**b**) Medium scale　　　　　　(**c**) Small scale

Figure 1. Typical pictures of smoke at different scale sizes: (**a**) large percentage of target; (**b**) medium percentage of target; (**c**) small percentage of target.

2.2. Model Structure of YOLOv8

The YOLOv8 model is divided into five scales (YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, YOLOv8x) of increasing size according to the network's depth and the feature map's width. The architecture design comprises three primary components, as visually depicted in Figure 2: Backbone, Neck and Head.

Backbone mainly consists of three modules: CBS (Convolution Block with SiLU activation), C2f (Cross-Stage Partial architecture with bottlenecks) and SPPF (Spatial Pyramid Pooling Fusion). CBS executes essential convolutional operations, incorporating 2D convolution, BatchNorm and SiLU activation functions. C2f adopts the innovative CSP (Cross-Stage Partial) architecture, which includes one split, two standard convolution layers and multiple bottleneck modules for feature extraction. SPPF module utilizes a collection of smaller pooling kernels that replace a solitary large pooling kernel. The fusion of features is accomplished by juxtaposing the feature map that has not undergone pooling with the feature maps produced after successive pooling operations. This method further optimizes the operational speed while preserving the potential for feature fusion across a wide receptive field. Neck is located between the Backbone and the output end. The proposed model incorporates the architectural design principles of both the FPN (Feature Pyramid Network) and PAN (Pyramid Attention Network), which consists of CBS, Upsample, Concat and C2f without shortcuts. The CBS within Neck is primarily responsible for extracting high-level semantic information via downsampling. C2f concentrates on extracting texture features from images, paying less attention to location and detail information

while extracting features and instead prioritizing semantic information. Extracted texture features are subsequently integrated with other features and channeled to the prediction layer, thus enhancing the ability of network feature fusion. Head uses the previously acquired features to make predictions and treats the classification task separately from the detection task. For the classification task, the cross-entropy loss function (BCE Loss) is still used to ensure the accurate classification of different categories. For the detection task, DFL (Distribution Focal Loss) and CIoU (Complete Intersection over Union) are adopted. DFL can better deal with the problem of class imbalance, which is helpful to improve accuracy. Simultaneously, the CIoU loss function effectively optimizes the position prediction of the detection boxes, further enhancing the performance of the detection and the ability of network feature fusion.
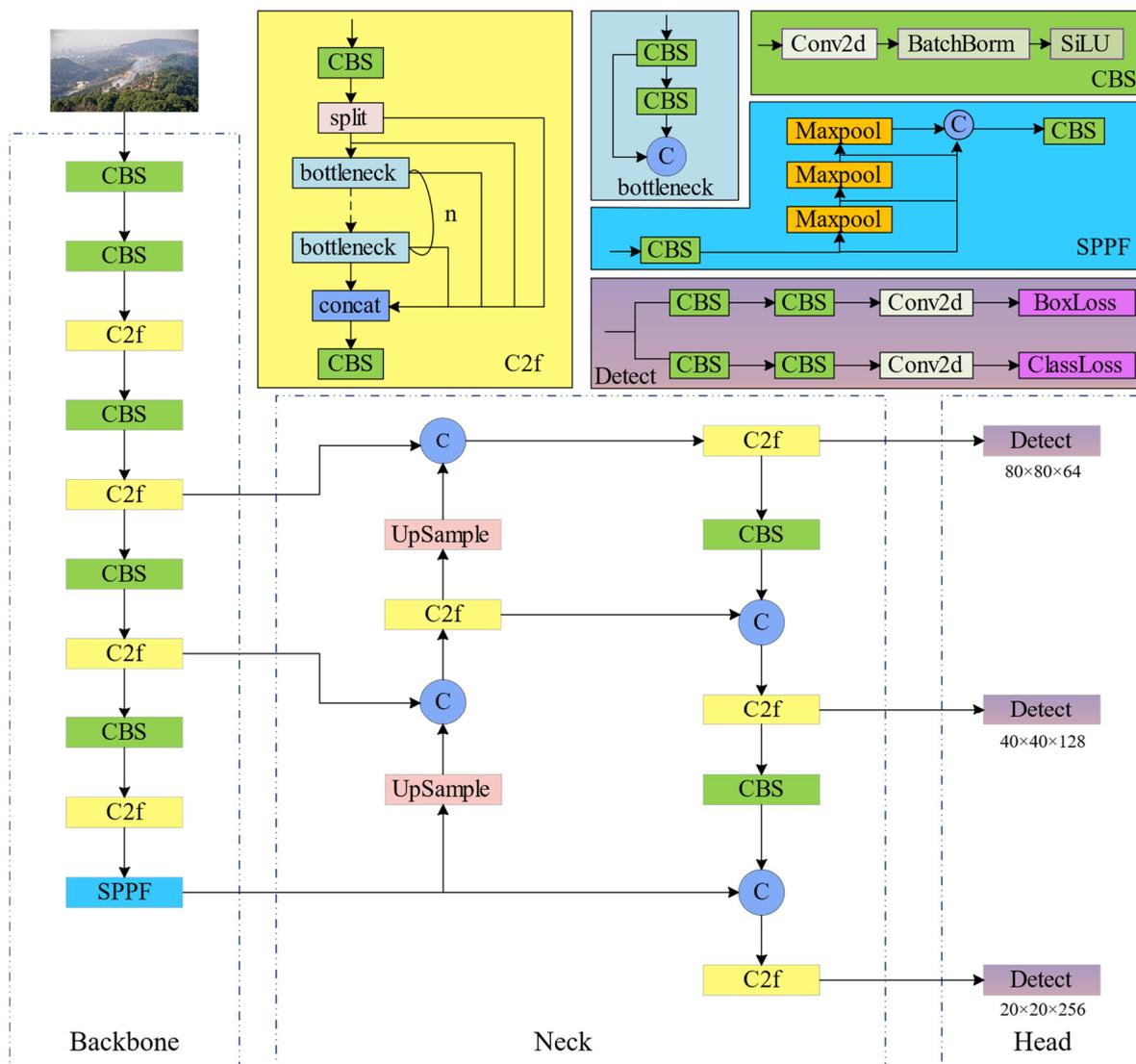

Figure 2. Structure of YOLOv8.

2.3. Space-to-Depth Convolution

Conventional CNN typically uses pooling layers and strided convolutions to decrease the spatial dimensions of images and extract features. However, for images with low resolution or small objects, these traditional methods may result in information loss or blurring, which can lead to a decline in overall performance. In this regard, Sunkara et al. [24] proposed a new CNN building block that eliminates the need for strided convolution and pooling. This building block is designed to handle low-resolution images

or small-size objects more effectively. As shown in Figure 3, this innovative building is named SPD-Conv. It is composed of a space-to-depth layer and a non-strided convolution layer, which downsamples the feature map and retains all the information of the channel dimension. In detail, a feature map denoted X has dimensions $S \times S \times C_1$. It initiates by segmenting this feature map N times along the channel dimension of the $S \times S$ plane (N = 2 in the provided illustration). Each resultant subfeature map stemming from this division retains the same number of channels, while the length and width shrink to S/N. These subfeature maps are merged along the channel dimension. The merged subfeature map X_1 undergoes convolution through a non-strided convolution layer to obtain a feature map X_2 that retains as much feature information as possible.
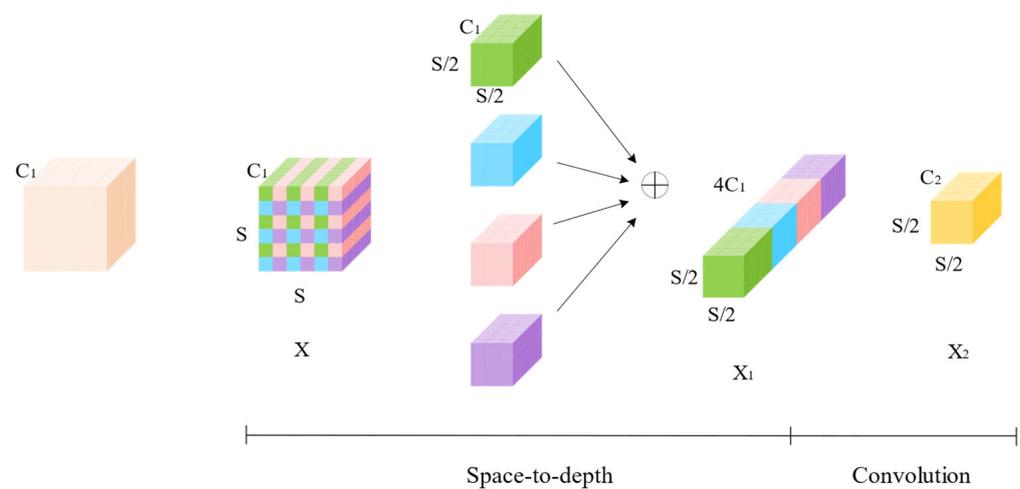

Figure 3. Structure of SPD-Conv module. X, X_1 and X_2 denote different stages of the feature maps; S represents the length and width of the original feature map; C indicates the number of channels in the feature map, representing the depth of the feature map; N denotes the number of segmentation operations.

2.4. GSConv

Conventional convolution requires extensive computation when dealing with large amounts of data, which can potentially result in slower training speeds. In order to overcome this problem, researchers put forward lightweight convolution [25,26]. By limiting the depth and width of the model, computational demands can be controlled, rendering it more suitable for deployment. Many lightweight networks employ a considerable amount of DSC (Depthwise Separable Convolution), but this often results in decreased detection accuracy. As a solution to this problem, Li et al. [27] proposed a new lightweight convolution known as GSConv. GSConv combines depth-separable convolution with standard convolution, uses depth-separable convolution to reduce the computational complexity and, at the same time, mitigates the low recognition accuracy caused by the poor feature extraction and fusion ability of depth-separable convolution via standard convolution. In forest fire detection, the model is too large to be deployed. To achieve comparable detection accuracy while reducing network computation, the GSConv module was used in this paper. The module's structure, depicted in Figure 4, encompassed Conv, DWConv, Concat and shuffle operations. The input feature map obtains the information that the number of channels is half of the number of output channels via a standard convolution. In the case where the number of channels remains unchanged via the depth-separable convolution, the two channels are then spliced together to recover the original number of channels. The final result is output through the Shuffle module.
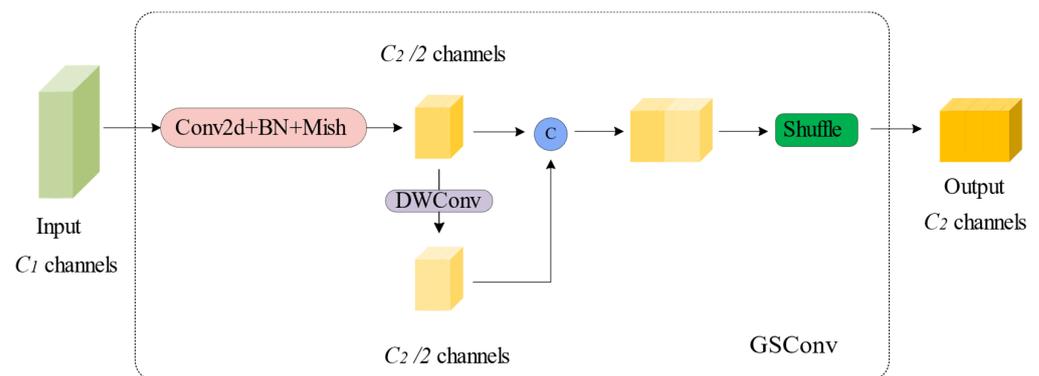
Figure 4. Structure of GSConv module. C refers to channels; the box marked in pink here refers to Conv mentioned in the text; the DWConv means the DSC operation; purple circle with the letter C refers to Concat operation.

2.5. ForestFireDetector

In addition to focusing on the accuracy in practical applications, the computational cost is also worth considering. So, this paper selects YOLOv8n as the base model, which downsamples using strided convolution that is susceptible to information loss. With the aim of enhancing the model's detection ability, we consider increasing the channel depth to enhance feature extraction. This way, during downsample, it can effectively reduce the loss of fine-grained information and mitigate the impact of insufficient feature learning, while preserving more discriminative information. To achieve this, we modify the structure. First, we modify the feature extraction part. We introduce SPD-Conv module in the feature extraction phase and add it before each C2f module. The space-to-depth layer divides the input feature map into spatial blocks and then rearranges the pixels in each block to form a new feature map with a smaller size, enabling downsampling. Second, we modify the feature fusion part. We replace the standard convolution of the Neck with GSConv. Applying it to the whole model would greatly deepen the network and increase the inference time, so it is only used in Neck. Based on GSConv, the structure of the GS bottleneck module uses 1×1 convolution to reduce the number of channels by half, and then 3×3 convolution doubles the channel count. Notably, the bottleneck in the Neck does not use a shortcut. After that, changing the activation function means changing CBS module to CBM module. The bottleneck in C2f module is replaced with GSbottleneck, and three GSbottlenecks are connected in series, which enhances accuracy while greatly reducing the number of parameters. Finally, we modify the feature processing part. There are only three detection scales in the YOLOv8 network structure, which predicts small, medium and large objects after downsampling the input image with a size of 640×640 sequentially. However, the smoke dataset contains smaller detection targets; using three detection heads results in insufficient utilization of shallow feature information, which causes a decrease in detection accuracy. Therefore, in this paper, we add a detection head for detecting smaller targets, building upon the foundation of the initial prototype. The improved model is depicted in Figure 5. We increase the channel depth using two convolutional structures that efficiently downsample to preserve feature details within the channel dimension while maintaining a balance between computational efficiency and performance enhancement.
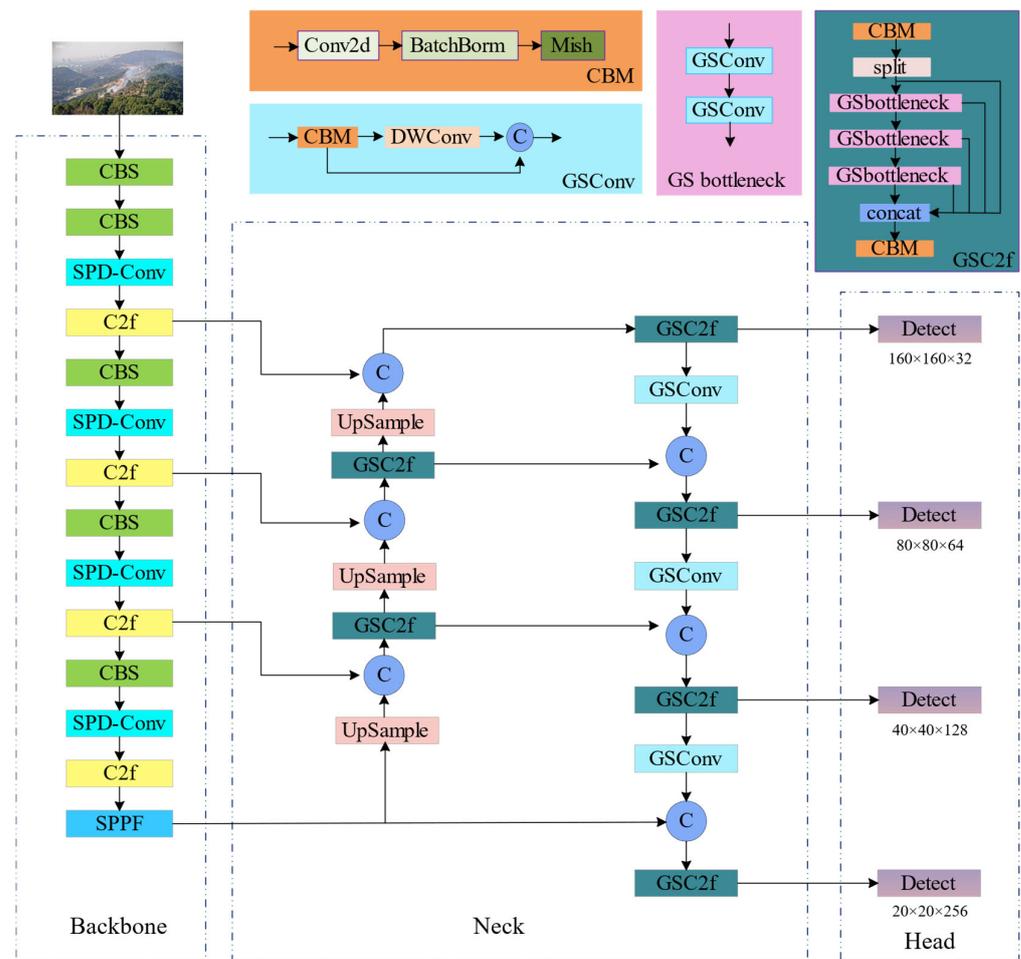
Figure 5. Structure of ForestFireDetector model.

3. Results

3.1. Training

The experimental environment particulars are outlined in Table 1. Training parameters of the improved model are set, as shown in Table 2. According to past experience, the dataset is divided into training, validation and test sets in an 8:1:1 ratio.

Table 1. Experimental environments.

Experimental Environment	Details
Programming language	Python 3.8
Operating system	Linux
Deep learning framework	Pytorch 1.11
GPU	NVIDA Tesla V100 PCle
Acceleration Tool	CUDA11.3

Table 2. Training parameters setting for ForestFireDetecor model.

Training Parameters	Details
Epochs	200
Batchsize	32
Img-size	640×640
Optimization	SGD
Initial learning rate	0.01

3.2. Model Evaluation

This experiment uses Precision (*P*), Recall (*R*) and mAP50 as evaluation metrics to assess the model performance. The calculation equations are as follows:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} \int_0^1 P(r) dr \tag{3}$$

In the above equation, TP denotes the count of positive samples correctly identified, FP refers to the count of positive samples with false positives, and FN represents the count of missing positive samples. Taking the experimental dataset as an example, Precision denotes the accuracy of correctly predicted smoke, with higher Precision indicating fewer misdetections. Recall indicates the coverage rate of correctly predicted smoke, and the greater Recall indicates that less smoke is missed. mAP50 refers to the mean average precision at IoU = 0.5.

3.3. Detection Performance and Analysis

The convergence process of Precision, Recall and mAP50 during the training of the improved model is shown in Figure 6. In the initial 50 epochs, there is a rapid increase in Precision, Recall and mAP50 metrics, indicating significant progress in the model's performance. Subsequently, substantial accuracy levels are attained following 100 epochs of training. Remarkably, the model reaches near-optimal performance around 150 epochs, underscoring its efficacy and robustness in terms of Precision, Recall and mAP50 metrics.
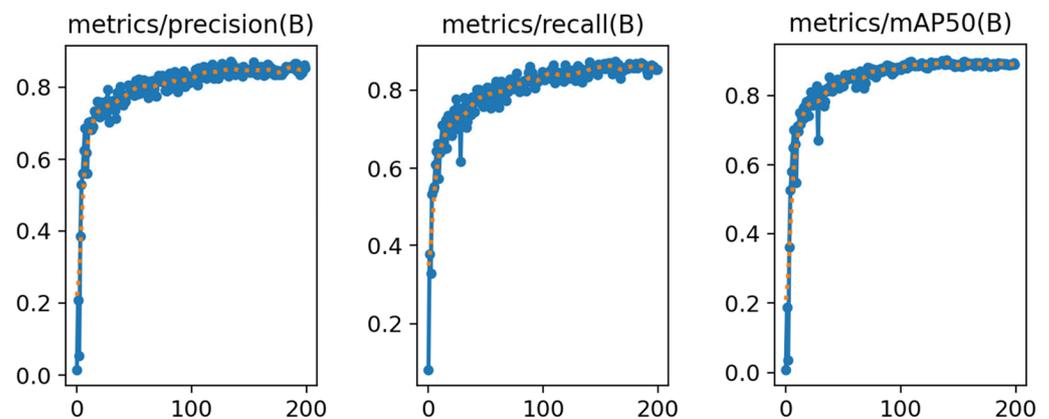

Figure 6. The convergence process of training.

To comprehensively validate the effectiveness of this model in detecting forest fire smoke, a comparative analysis was conducted against other mainstream models, including Faster R-CNN, SSD and YOLOv5s (Table 3). Among the different models, we are more concerned about which model has a lower missed detection rate. The results show that, in terms of Recall, our model outperforms the others except for YOLOv5s. However, the distinct advantage of our model lies in achieving a higher mAP compared to other models, all while maintaining the least number of parameters. This unique combination positions our model as particularly well suited for early forest fire detection.

Table 3. Comparison of experimental results.

Model	Precision	Recall	mAP50	Params	Flops
Faster RCNN [21]	0.388	0.728	0.801	136.69	369.72
SSD [16]	0.801	0.768	0.832	23.61	62.7
YOLOv5s [28]	0.853	0.879	0.874	7.06	16.5
YOLOv8n [29]	0.85	0.796	0.869	3.01	8.2
FireDetector (Ours)	0.852	0.871	0.902	2.87	14.8

The original YOLOv8 detection model is trained using the smoke dataset and then evaluated with the validation set, resulting in an mAP50 of 0.869. The results indicate that as a more advanced detection model, it performs well for smoke detection. However, there is still potential for further enhancements. To evaluate the efficacy of the enhancements based on the YOLOv8 model, each improvement is tested individually, and the corresponding outcomes of the experiments are detailed in Table 4. Introducing four SPD-Conv modules into the Backbone network yielded a 2.2% improvement in mAP50, although this came at the cost of heightened computational demands. Just replacing the convolution in Neck yielded a 2.5% enhancement while simultaneously reducing computational requirements. Even the addition of only the detection head led to accuracy improvements, although the accuracy improvement is not as great as the first two points. Building upon these singular enhancements, the effects of pairwise combinations were examined. Notably, the synergy of SPD-Conv and GSConv garnered remarkable accuracy improvements without incurring an unwarranted increase in parameters. The experimental results demonstrated that the combination of these three enhancements positively impacted the model's detection performance. Consequently, these refinements effectively bolstered the model's detection capabilities, further establishing its proficiency in detecting smoke.

Table 4. Results of ablation experiments.

YOLOv8	SPD-Conv	GSConv	Head	Precision	Recall	mAP50	Params	FLOPs
√				0.85	0.796	0.869	3011043	8.2
√	√			0.848	0.844	0.891	3272163	11.7
√		√		0.86	0.836	0.894	**2408131**	**7.0**
√			√	0.859	0.821	0.889	2926692	12.4
√	√	√		**0.883**	0.838	**0.91**	2882979	11.0
√	√		√	0.846	0.85	0.893	3187812	15.9
√		√	√	0.839	0.861	0.905	2519236	11.3
√	√	√	√	0.852	**0.871**	0.902	2780356	14.8

From these experiments, it is evident that our improved YOLOv8 model demonstrates its effectiveness in detecting small and inconspicuous smoke, surpassing other models in terms of accuracy at various scales. The detection performance, both pre- and post-enhancement, is displayed in Figure 7. As shown in Figure 7, the outcomes of the enhanced model show no instances of false or missed detections. Furthermore, the bounding boxes demonstrate a remarkable level of precision as they effectively encapsulate the entirety of the smoke objects.

(**a**) Detection results of medium scales.

(**b**) Detection results of small scales.

(**c**) Detection results of missed smoke.

(**d**) Detection results of false smoke.

Figure 7. Detection effect before (left) and after (right) the improvement: (**a**,**b**) smoke at different percentage detection results; (**c**,**d**) missed and false detection results.

4. Discussion

Forest fires have a far-reaching negative impact on ecosystems, human health and social economics. Effective fire prevention and control are pivotal to maintaining the ecological balance of the Earth, achieving sustainable development for human society, and ensuring the normal functioning of ecosystems. Smoke, as an important feature of early fire, should be given more attention. Taking appropriate measures using relevant technologies and methods within the domain of target detection holds significance in practice. Although detection technologies are getting better, difficulties still exist in detecting smoke. Therefore, further research on smoke detection is needed. Through experimentation, we found that YOLOv8 has a better detection effect, but there is still potential for further enhancements. To

this end, the following improvements were made: First of all, the presence of images with small targets and low resolution in the dataset could decrease the detection performance. In order to solve this problem, the channel depth is increased by incorporating the SPD-Conv module into the Backbone to mitigate the negative impact of the loss of fine-grained information and insufficient learning features and to retain more discriminative information, thus improving the detection effect on the smaller smoke. The essence of this approach lies in its ability to effectively downsample while safeguarding feature details within the channel dimension, thereby catering to the nuances of low-resolution images and small objects. The experimental results show that mAP is improved by 2.2%. Additionally, YOLOv8′s large downsampling multiple makes it challenging to learn feature information for small targets from the deep feature map. Addressing this, we added a small target detection layer to identify features in both the shallow and deep feature maps following their fusion. By incorporating a small target detection layer, the network can focus on identifying tiny targets, leading to improved detection capability; mAP is improved by 2%. However, with the increase in the detection layer, despite enhanced detection capability, an increase in computational load ensued. Hence, as a remedy, we substituted the convolution of the Neck with GSConv, expanding channel depth by standard convolution and depth-separable convolution, which reduces computational effort while improving the mAP by 2.5%. We struck a balance between computational efficiency and performance enhancement, rendering it particularly valuable in scenarios like forest fire detection, where both speed and precision are vital.

In order to achieve a better application effect, future research can be carried out in the following areas: On the one hand, the model can be further optimized. For the current model, continuous optimization and improvement are carried out to improve its training effect and detection capabilities. Potential optimization directions include modifying the network structure, improving loss function, optimizing the hyperparameters and so on. On the other hand, we can also consider using a larger dataset for training to increase the model's capacity for generalization. Moreover, studying the detection model suitable for both forest fire and smoke. At present, research mainly focuses on the detection of smoke, but in the actual situation, it is necessary to consider both the detection of fire and smoke simultaneously. Therefore, a comprehensive model suitable for detecting forest fires and smoke can be explored. This may involve the method of fusing various sensor data or image information to improve the detection performance of the model in a complex environment.

5. Conclusions

The proposed method described in this paper achieves the detection of forest fire smoke. The results from the experiments demonstrate the model's capability to precisely detect and identify smoke images in the forest. The introduction of the space-to-depth layer significantly improves the ability to capture contextual smoke information. This facilitates the model to capture feature information of forest fire smoke, which enhances the model's feature extraction capabilities and augments the learning capabilities of the network. In addition, in the GSConv module, when depth-separable convolution is used, the depth of the channel is increased while its own properties are used to reduce the parameters of the model. Overall, significant improvements are made in fine-grained feature learning by expanding channel depth, resulting in better detection of forest fire smoke. In comparison to other models, the model proposed in this paper has a good detection and recognition effect, and its performance advantage is remarkable.

In conclusion, this study not only contributes a refined methodology for forest fire smoke detection but also demonstrates the potency of the proposed model using rigorous experimentation. The augmentation of fine-grained learning and feature extraction via the proposed improvements showcases the model's substantial advancement over existing models. This model holds the potential to render an effective contribution to the field of early smoke detection and prevention.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The dataset and code used during this study are not publicly available due to their particularity.

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

References

1. Chen, S.; Chen, J.; Jiang, C.; Yao, R.T.; Xue, J.; Bai, Y.; Wang, S. Trends in research on forest ecosystem services in the most recent 20 years: A bibliometric analysis. *Forests* **2022**, *13*, 1087. [CrossRef]
2. Wang, L.R.; Huang, T.L.; Li, Y. Classification and research status of forest natural disasters. *For. Sci. Technol. Commun.* **2022**, *8*, 8–13.
3. Wang, M.Y.; Xu, Y.; Zhao, M.W. Analysis of the spatial and temporal distribution pattern and causes of forest fires in China in the past 10 years. *Agric. Sci. Technol. Commun.* **2021**, *10*, 201–204.
4. Statistical Data-Emergency Management Department of People's Republic of China (PRC). The Emergency Management Department Released the National Natural Disasters in the First Half of 2023. Available online: https://www.mem.gov.cn/gk/tjsj/ (accessed on 31 August 2023).
5. Krüll, W.; Tobera, R.; Willms, I.; Essen, H.; von Wahl, N. Early Forest Fire Detection and Verification Using Optical Smoke, Gas and Microwave Sensors. *Procedia Eng.* **2012**, *45*, 584–594. [CrossRef]
6. Von Wahl, N.; Heinen, S.; Essen, H.; Kruell, W.; Tobera, R.; Willms, I. An Integrated Approach for Early Forest Fire Detection and Verification Using Optical Smoke, Gas and Microwave Sensors. *WIT Trans. Ecol. Environ.* **2010**, *137*, 97–106.
7. Yang, N.; Wang, Z.; Wang, S. Computer Image Recognition Technology and Application Analysis. In *Proceedings of the IOP Conference Series Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2021; Volume 769, p. 032065.
8. Li, M.N.; Zhang, Y.M.; Mu, L.X.; Xin, J.; Yu, Z.Q.; Liu, H.; Xie, G. Early forest fire detection based on deep learning. In Proceedings of the 2021 3rd International Conference on Industrial Artificial Intelligence, Shenyang, China, 8–11 November 2021; pp. 1–5.
9. Ya'acob, N.; Najib, M.S.M.; Tajudin, N.; Yusof, A.L.; Kassim, M. Image Processing Based Forest Fire Detection Using Infrared Camera. *J. Phys. Conf. Ser.* **2021**, *1768*, 012014. [CrossRef]
10. Miao, Z.W.; Lu, Z.N.; Wang, J.L.; Wang, Y. Research on fire detection based on vision. *For. Eng.* **2022**, *38*, 86–92.
11. Ding, X.; Gao, J. A New Intelligent Fire Color Space Approach for Forest Fire Detection. *J. Intell. Fuzzy Syst.* **2022**, *42*, 5265–5281.
12. Huang, X.Y.; Lin, S.R.; Liu, N.A. A Review of Smoldering Wildfire: Research Advances and Prospects. *J. Eng. Thermophys.* **2021**, *42*, 512–528.
13. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
14. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
15. 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. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475.
16. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; 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; pp. 21–37.
17. Zhu, M.X.; Liu, Z.Q.; Zhang, X.; Li, W.J.; Su, J.X. Review of Research on Video-Based Smoke Detection Algorithms. *Comput. Eng. Appl.* **2022**, *58*, 16–26.
18. Wu, S.; Zhang, L. Using Popular Object Detection Methods for Real Time Forest Fire Detection. In Proceedings of the 2018 11th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 8–9 December 2018; pp. 280–284.
19. Al-Smadi, Y.; Alauthman, M.; Al-Qerem, A.; Aldweesh, A.; Quaddoura, R.; Aburub, F.; Mansour, K.; Alhmiedat, T. Early Wildfire Smoke Detection Using Different YOLO Models. *Machines* **2023**, *11*, 246. [CrossRef]
20. Wan, Z.; Zhuo, Y.; Jiang, H.; Tao, J.; Qian, H.; Xiang, W.; Qian, Y. Fire Detection from Images Based on Single Shot MultiBox Detector. In *Advances in Intelligent Systems and Computing, Proceedings of the The 10th International Conference on Computer Engineering and Networks, Singapore, 6 October 2020*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 302–313.

21. Ren, S.; He, K.; 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]
22. Zhao, L.; Zhi, L.; Zhao, C.; Zheng, W. Fire-YOLO: A Small Target Object Detection Method for Fire Inspection. *Sustainability* **2022**, *14*, 4030. [CrossRef]
23. Rahman, E.U.; Khan, M.A.; Algarni, F.; Zhang, Y.; Irfan Uddin, M.; Ullah, I.; Ahmad, H.I. Computer Vision-Based Wildfire Smoke Detection Using UAVs. *Math. Probl. Eng.* **2021**, *2021*, 5594899. [CrossRef]
24. Sunkara, R.; Luo, T. No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Grenoble, France, 7 August 2022; pp. 443–459.
25. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
26. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. *arXiv* **2017**. [CrossRef]
27. 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.
28. GitHub. Ultralytics/Yolov5: YOLOv5 in PyTorch > ONNX > CoreML > TFLite. Available online: https://github.com/ultralytics/yolov5 (accessed on 4 September 2023).
29. GitHub. Ultralytics/Ultralytics: NEW—YOLOv8 in PyTorch > ONNX > OpenVINO > CoreML > TFLite. Available online: https://github.com/ultralytics/ultralytics (accessed on 4 September 2023).