FOREST FIRE DETECTION BASED ON IMPROVED YOLOV7 MODELING

YANG, Q.[1*] – ZHANG, T.[1] – TONG, X.[2] – HU, L. H.[1]

[1]*Department of artificial intelligence and big data, Yibin University, Yibin 644000, China*

[2]*School of Information Engineering, Chengdu University of Technology, Chengdu 61007, China*

[*]*Corresponding author*
e-mail: scyangqiang@163.com; phone: +86-159-0839-2479

(Received 8th Jan 2024; accepted 3rd May 2024)

Abstract. An improved forest fire detection algorithm based on the You Only Look Once v7 (YOLOv7) is proposed in this paper to address the poor accuracy and speed of traditional forest fire detection. In order to guide this model to better focus on the features of forest smoke and fire, this paper introduces the Convolutional Block Attention Module (CBAM) channel attention mechanism and spatial attention mechanism in the backbone network and the head of the YOLOv7 model. It uses the SCYLLA-IoU (SIoU) loss function to replace the Complete-IoU (CIoU) loss function in order to improve the regression accuracy and convergence speed. Experiments show that the improved YOLOv7 algorithm achieves 78.8%, 65.4%, and 70.8% accuracy, recall, and average accuracy, respectively. Compared with the original algorithm, the accuracy, recall, and average accuracy are improved by 4.7% ,1.6% and 5.3%. The algorithm reduces missed detections, improves the accuracy and practicality of forest fire detection, and has good robustness in different complex scenarios.
Keywords: *fire detection, deep learning, YOLO, SIou, attention mechanism*

Introduction

Forest fires can jeopardize property and life and can cause extensive damage to the environment (Alkhatib et al., 2014). Traditional approaches such as deploying fire and smoke detection sensors (Chen et al., 2003), manual inspection, image processing (Chen et al., 2004), etc., exhibit problems such as the lack of timeliness, the possibility of misjudgments, omissions, etc., and can be a waste of resources (Barmpoutis et al., 2020; Kim et al., 2023; Abdusalomov et al., 2023; Chen et al., 2023). Deep learning algorithms are developing rapidly in the area of image recognition and detection. Fire detection technology based on deep learning has the advantages of fast detection speed, high accuracy, low cost, and wide range of applications (Jiao et al., 2019; Guede-Fernández et al., 2021; Seydi et al., 2022; Bahhar et al., 2023). We proposed an improved target detection method based on the YOLOv7 (WANG et al., 2022) model. The model can directly interact with cameras connected to the computer in the forest, capture video streams in real time, and extract the current frame for fireworks detection. This direct interaction with hardware enables the model to respond instantly to environmental changes, making it very suitable for deployment in places that require real-time fireworks monitoring. The goal of this study is to address the shortcomings of the traditional fire detection model and the lack of smoke detection. It has a significant impact on life and ecological environment protection.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136

Basic Structure of the YOLOv7 Model

The YOLOv7 network model is mainly composed of four parts: the input, the backbone network, the neck, and the head. The network structure is shown in *Figure 1* and the structure of each module is shown in *Figure 2*.

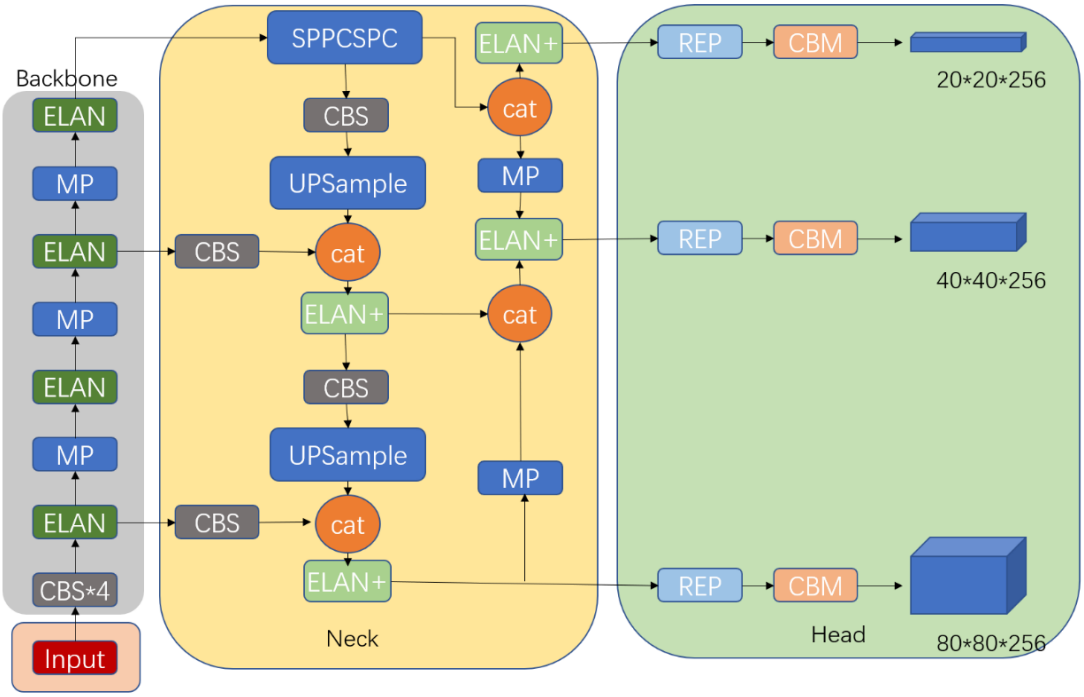

Figure 1. *YOLOv7 network structure*

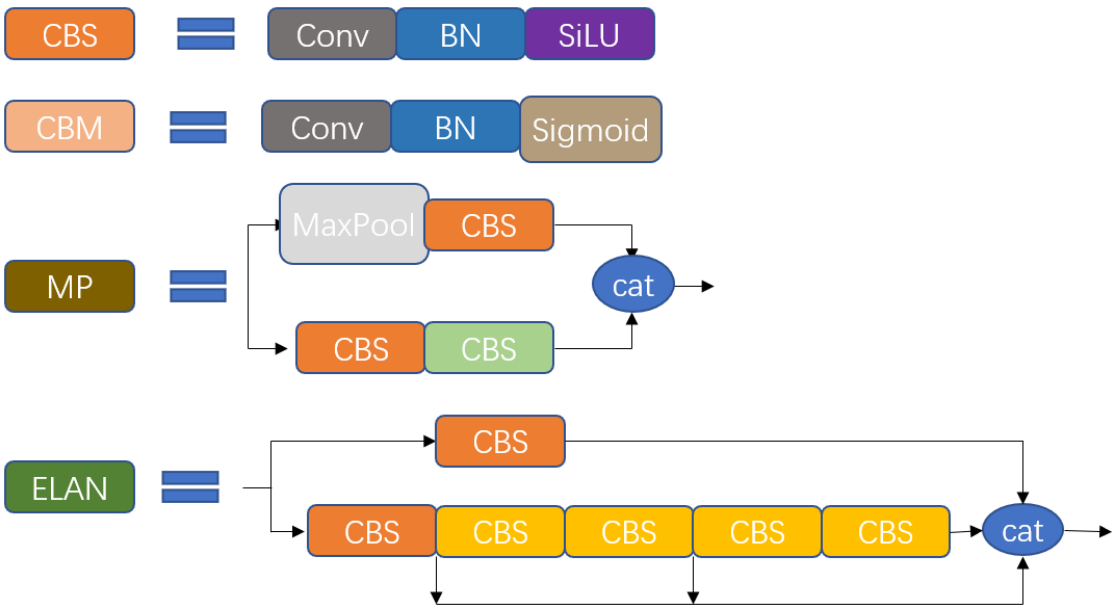

Figure 2. *Structure of each module*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136

Input

The input part uses Mosaic data enhancement, adaptive anchor frame calculation, and adaptive image scaling to preprocess the input image.

Backbone Network

The backbone network consists of several Conv + BN + Silu (CBS) modules, Efficient Layer Aggregation Networks (ELAN) modules, and MaxPool (MP) modules together. The Conv denotes the convolution layer, the BN denotes the batch normalization layer, and Silu and Sigmoid denote the activation function. The CBS module consists of a convolutional layer, a BN layer, and a SiLU activation function. The ELAN layer consists of seven CBS modules, which can effectively learn and quickly converge by controlling the shortest and longest gradient paths, and also improves the robustness of the model. The MP layer consists of three CBS modules and a maximum pooling layer. Both parts downsample the feature map and change the number of channels, and finally fuse their results which enhances the feature extraction capability.

Neck

The neck part adopts the module of Spatial Pyramid Pooling modification module (SPPCSPC) and ELAN+. The SPPCSPC module has four different scales to distinguish between large and small targets and increases the receptive field by max pooling to adapt to images with different resolutions. The UPSample module performs upsampling by nearest neighbor interpolation. The difference between ELAN+ and ELAN is the number of outputs selected.

Head

The head part uses REP and Conv + BN + Sigmoid (CBM) modules. The REP module outputs three feature maps with different scale sizes by extracting features and outputs three prediction results through the REP and convolutional layer. The CBM layer is similar to the CBS layer; the difference is that the activation function is replaced with sigmoid.

The Improved YOLOv7 Network Model

CBAM Attention Mechanism

By observing the flame and smoke images in the dataset, it is found that there are significant differences in the distribution of flame and smoke positions in some images, and they are greatly affected by different conditions such as weather and lighting. Therefore, a new lightweight attention model is introduced. An attention mechanism is a special neural network structure that allows more important regions to receive more attention when the backbone neural network learns features. CBAM (Woo et al., 2018) (Convolutional Block Attention Module) contains two independent sub-modules, the Channel Attention Module (CAM) and the Spatial Attention Module (SAM), which perform the fusion of attention features in the channel and spatial dimensions, respectively. This not only saves parameters and computational power, but also ensures that it can be integrated into existing network architectures as a plug and play module.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136

Embedding the CBAM attention module into the YOLOv7 network is beneficial for solving the problem of "no attention preference" in the original network.

SIoU Loss Function

The YOLOv7 loss function involves three parts: confidence loss, classification loss, and localization loss, using the CIoU (Zhaohui et al., 2020) loss function for bounding box regression. CIoU takes into account the overlap area, centroid distance, and aspect ratio, and the function is given in *Eq 1-4*.

$$L_{ciou} = 1 - IOU + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v \quad \text{(Eq.1)}$$

$$IOU = \frac{A \cap B}{A \cup B} \quad \text{(Eq.2)}$$

$$\alpha = \frac{v}{1-IOU+v} \quad \text{(Eq.3)}$$

$$v = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2 \quad \text{(Eq.4)}$$

where IoU denotes the intersection and concurrency ratio of the ground truth bounding box and the predicted bounding box, and A and B denote the ground truth bounding box and the predicted bounding box; b and b^{gt} represent the centroids of the predicted bounding box B and the ground truth bounding box B^{gt}, respectively, and ρ^2 denotes the Euclidean distance; c is the diagonal distance of the minimum outer rectangle of the predicted bounding box and the ground truth bounding box; w^{gt}, h^{gt}, w, and h stand for the width and height of the ground truth bounding box and for the predicted bounding box, respectively; v denotes the consistency between the ground truth bounding box and the predicted bounding box.

Compared with CIoU, SIoU (Gevorgyan, 2022) also takes into account the angle, distance, and shape factors which accelerates the convergence speed and optimizes the regression performance.SIoU consists of the angle loss L_{angle}, the distance loss $L_{distance}$, the shape loss L_{shape}, and the IoU loss, which are defined as shown in *Eq 5-10*; the meanings of the parameters are shown in *Figure 3*.

$$L_{siou} = 1 - IOU + \frac{L_{distance}+L_{shape}}{2} \quad \text{(Eq.5)}$$

$$L_{angle} = sin(2\alpha) \quad \text{(Eq.6)}$$

$$L_{distance} = 2 - e^{-\upsilon\rho_x} - e^{-\upsilon\rho_y} \quad \text{(Eq.7)}$$

$$\rho_x = (\frac{b_{cx}^{gt}-b_{cx}}{C_w})^2, \rho_y = (\frac{b_{cy}^{gt}-b_{cy}}{C_h})^2, \upsilon = 2 - L_{angle} \quad \text{(Eq.8)}$$

$$L_{shape} = (1 - e^{-\lambda_w})^\theta + (1 - e^{-\lambda_h})^\theta \quad \text{(Eq.9)}$$

$$\lambda_w = \frac{|w-w^{gt}|}{max\{w,w^{gt}\}}, \lambda_h = \frac{|h-h^{gt}|}{max\{h,h^{gt}\}} \quad \text{(Eq.10)}$$

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136

where $(b_{cx}^{gt}, b_{cy}^{gt})$ and (b_{cx}, b_{cy}) denote the coordinates of the center points of the ground truth bounding box and the predicted bounding box, respectively; w^{gt}, h^{gt}, w, and h are the width and height of the ground truth bounding box and the predicted bounding box, respectively; C_w and C_h denote the width and height of the smallest outer bounding rectangle of the predicted and ground truth bounding box, and α is the acute angle made by the line connecting the two center points and the horizontal direction. θ in *Eq 9* is the parameter that controls the attention to shape loss with the parameter range of Chen et al. (2004) and Abdusalomov et al. (2023).

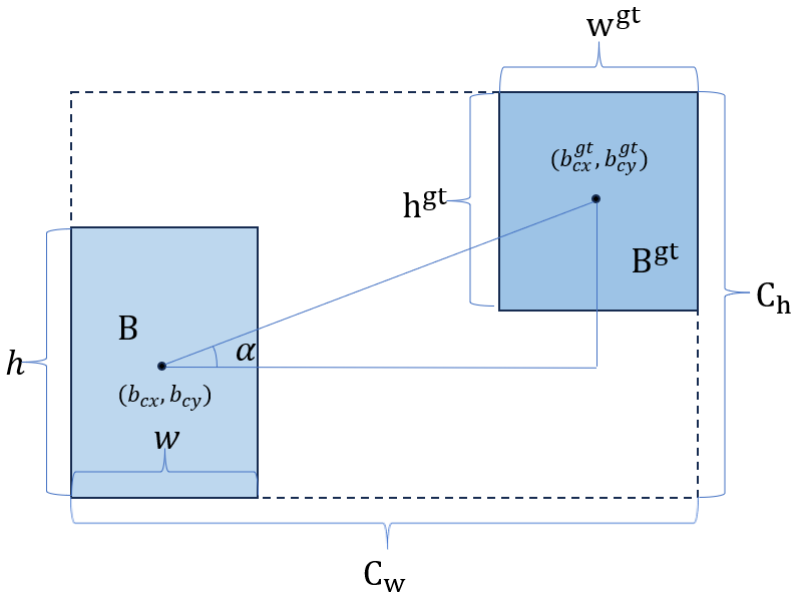

Figure 3. *The parameters*

As shown in *Figure 4*, it is the improved YOLOv7 network structure in this paper. Firstly, the original image is enhanced at the input; Then, by introducing CBAM attention mechanism into the backbone and neck of YOLOv7, the network's ability to focus on important features is effectively improved; Finally, the SIoU loss function is used for boundary box regression to accelerate the loss gradient convergence of the model.

Algorithm implementation

As stated previously, the proposed forest fire detection algorithm is based on the YOLOv7 network structure, introduces the CBAM attention mechanism, and uses the SIoU loss function. The specific implementation steps are as follows:

Step 1: Input the initial image F.

Step 2: The initial image is enhanced in the input part to obtain the image F1.

Step 3: In the backbone network section, MaxPool and multi-scale feature extraction were performed on F1, and attention mechanisms were fused to obtain F2. $F2 = \mathrm{F1} \otimes \mathrm{MP} \otimes \sum f \otimes CBAM$; MP is the MaxPool operation, $\sum f$ denotes feature fusion, and $CBAM$ is the fusion attention mechanism.

Step 4: In the neck part, features at different scales are obtained after multiple MaxPool operations, and after upsampling, the multi-scale features are fused to obtain F3. $F3 = \mathrm{F2} \otimes \mathrm{MP} \otimes UP \otimes \sum f$, and UP denotes upsampling.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136

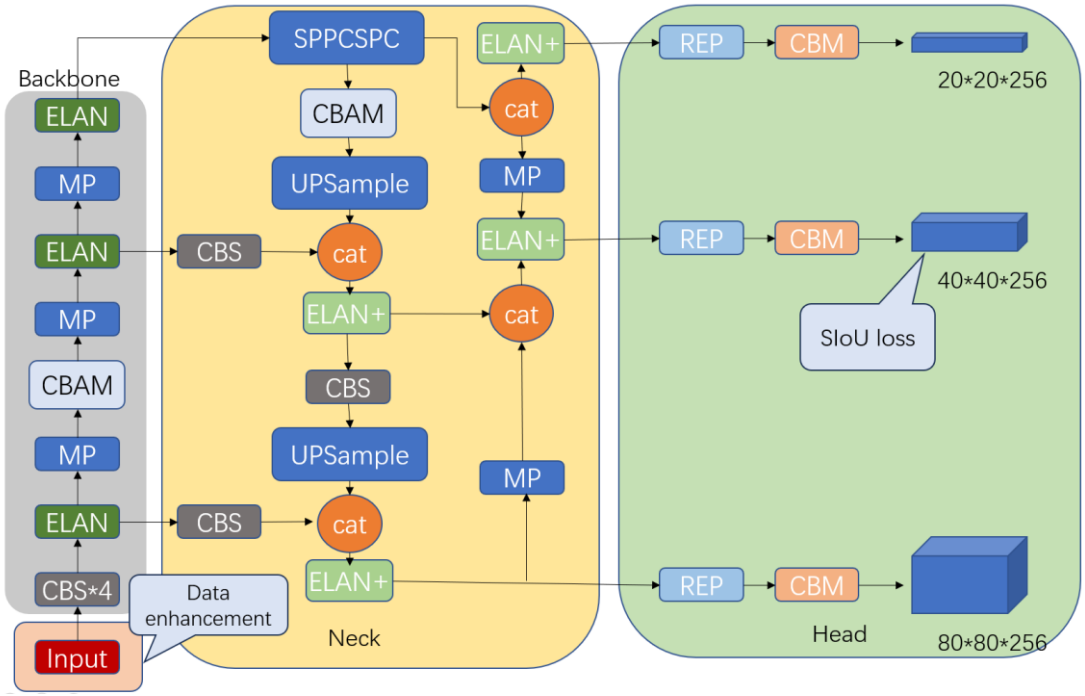

Figure 4. *The improved YOLOv7 network structure*

Step 5: In the head part, the large, medium, and small features obtained from the neck part are fused and the attention mechanism is used to improve the feature attention. This is fed into the detection head to output the final detection result, F4. $F4 = (F3_l + F3_m + F3_s) \otimes CBAM$; $F3_l$, $F3_m$, $F3_s$ denote the large, medium, and small scale features, respectively.

Step 6: Repeat steps one to five, keep iterating until convergence.

Experimentation and Analysis

Dataset Construction

The dataset in this article is sourced from online collection and video capture, such as Roboflow, Google Images, and Baidu Images. The dataset includes fireworks sample images, with an initial image of 3000 and a total of 1862 after sorting. The dataset is labeled using labeling tools, with categories of fire and smoke, and saved in the YOLO series txt format. It contains actual fire scenes under different weather conditions, different lighting, and different angles, which is a more complex picture environment and richer data than public datasets. It is divided into a training set of 1375 pictures, a validation set of 337 pictures, and a test set of 150 pictures. In order to improve the robustness and generalization of the model, the dataset contains 80 natural environment images without fire scenarios originated from Baidu Images as negative samples. Information about the source and number of labels in each category is shown in *Table 1*.

Experimental Configurations

The operating system of the experimental operating platform is Ubuntu 20.04.4 LTS, the graphics card is the NVIDIA GeForce RTX A5000, the processor is the 16 VCPU

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136

Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz, and the running memory is 42 GB. The learning framework for the experimentation is PyTorch 1.12, the experimental environment is Python 3.8, and the GPU acceleration software used is CUDA 11.3.

Table 1. *Information on the source and number of samples*

Sample type	Source	Quantity
Smoke	Roboflow	775
Fire	Google Images, Roboflow	1007
Fire-free natural environment	Baidu Images	80

Evaluation Indicators

The effectiveness of this YOLOv7 network model for the detection of forest fires is evaluated quantitively. This paper adopts precision, recall, and mean Average Precision to assess the model performance; the formulae are shown in *Eq. 11-13*.

$$P = \frac{TP}{TP+FP} \quad \text{(Eq.11)}$$

$$R = \frac{TP}{TP+FN} \quad \text{(Eq.12)}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP(i) \quad \text{(Eq.13)}$$

where P is the proportion of all predicted targets that are correct; TP (True Positive) is the number of positive samples that are detected correctly; FP (False Positive) is the number of negative samples that are detected as positive; FN (False Negative) is the number of backgrounds that are incorrectly detected as positive. R is the proportion of all labeled targets that are correctly detected; n denotes the number of categories; AP denotes the area of the P-R curve formed with R as the horizontal axis and P as the vertical axis to measure the model's detection performance on the category. mAP is obtained by de-averaging the AP calculated for each target category. When IoU is set to 0.5, the AP of all images in each category is calculated, and then all categories are averaged to obtain mAP@0.5.

Data Preprocessing

The resolution size of the dataset has a crucial role in the operation of the convolutional neural network model; if it is too large, it will take up a lot of computational and storage resources and burden the network model. Therefore, this paper proposes a preprocessing method that uses the MATLAB vision library to preprocess the input images, centering on flames and smoke, cropping, and normalizing, and adjusting the image resolution. The sample images are labeled for classification using the label annotation tool and the results are corrected.

Data Enhancement

This experiment crops and normalizes all image data to keep flames and smoke at the center of the image. In the dataset, there is a significant difference in the number of

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136

fireworks samples. If model training is conducted directly, it will have an impact on the recognition performance of smoke with a small number of samples. Therefore, in order to solve the problem of model performance degradation caused by imbalanced sample size, this article adopts various data augmentation methods, such as translation, flipping, rotation, increasing contrast, saturation, scaling, adding noise, multi image fusion, resolution adjustment, etc., to expand the number of pyrotechnic samples. The data enhancement effect is shown in *Figure 5*.

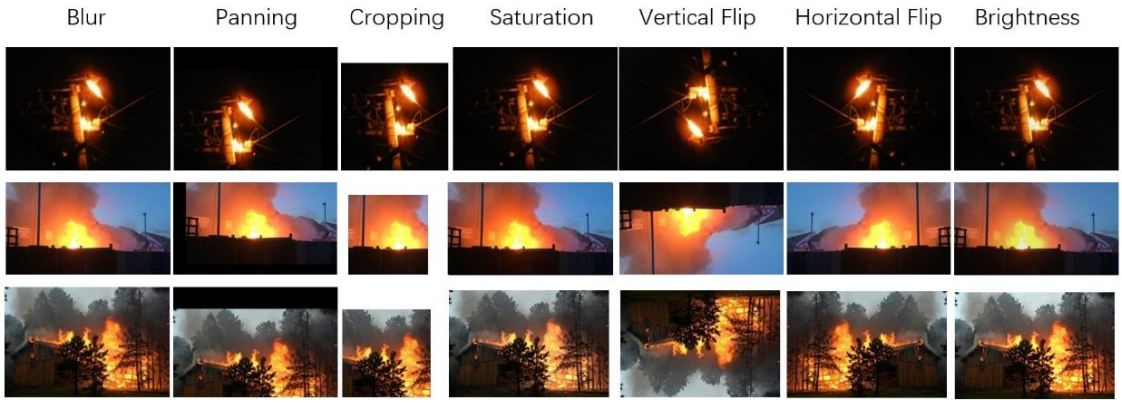

Figure 5. *Data enhancement*

Model Training

During the training phase, random flipping and random cropping, etc., are used as data enhancement techniques to avoid overfitting. To ensure model convergence, the network was trained for 300 epochs on an NVIDIA GeForce RTX A5000 GPU with a batch size of 16. The weight decay is 5e-4 with a momentum of 0.937. The input image was resized to 640×640.

For the model parameters, *epochs* are the number of training rounds; *batch-size* denotes the batch size; *lr* denotes the learning rate; *weight-decay* is the weight decay, which is a regularization technique that serves to inhibit the overfitting of the model as a way of improving the model's generalization; *size* is the size of the training images, and *momentum* is the momentum. After a large number of experiments and comparing the results of the experiments, the parameter settings were finally determined and are shown in *Table 2*.

Table 2. *Network parameter settings*

Parameter	Value
epochs	300
batch-size	16
lr	0.001
weight-decay	0.0005
size	640×640
momentum	0.937

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136

Analysis of Detection Performance

From *Figure 6*, it can be seen that (a) and (c) show the detection performance before improvement, while (b) and (d) show the detection performance after improvement. From (a) and (c) the integrity of smoke detection in forest fire by the model is poor, only part of the smoke is detected, and there are still cases of missing detection of small-scale flame. The precision rate and recall rate are 74.1% and 63.8%, which greatly affects the detection effect of fire. As an important index to evaluate the safety of the model, the recall rate after improvement has increased by 1.6%, and the precision rate has also increased by 4.7%. The recognition area and detection effect have been significantly improved. For the detection of columnar smoke, it is extremely easy to miss the detection before the improvement, and the alarm is not timely, which is very unfavorable for the prevention and control of forest fire and monitoring, and seriously affects the safety. The improved model has significantly improved the detection effect of columnar smoke, and also improved the safety and practicability.

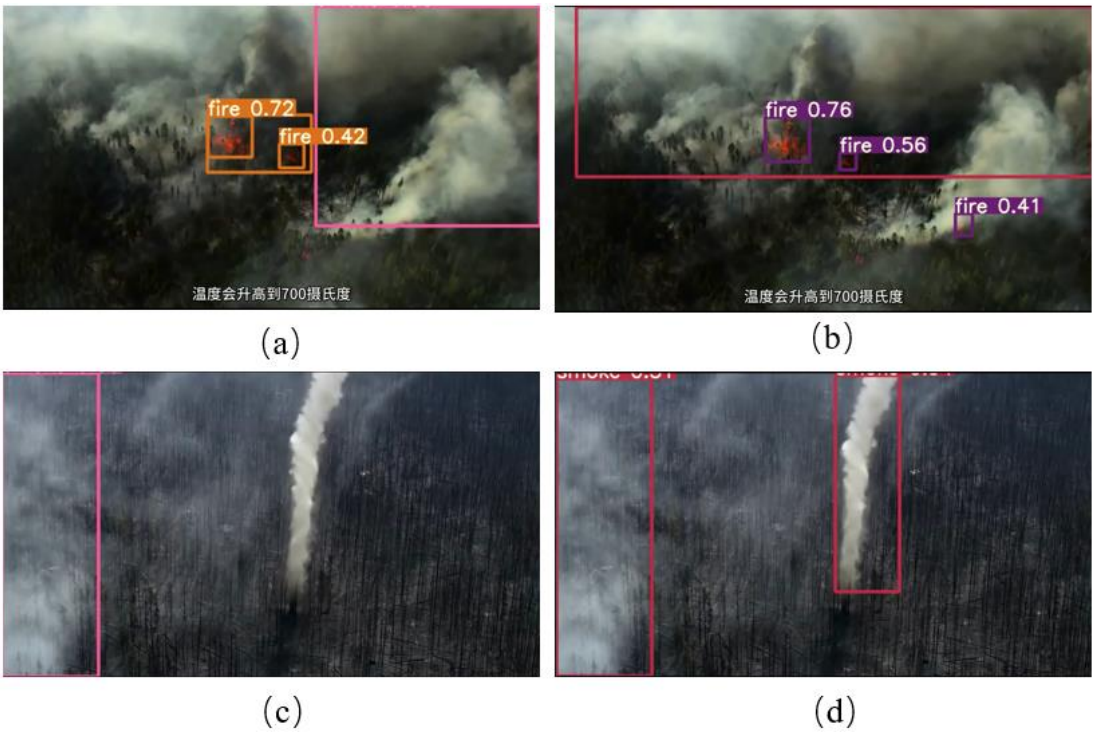

Figure 6. *Comparison of detection performance before and after model improvement*

Ablation Experiments

Due to the decreased model detection accuracy caused by the complexity of flame and smoke morphology and diverse environments, this paper employs data enhancement, the CBAM attention mechanism, the SIoU loss function, and other methods to optimize the fire detection model. Based on the improved YOLOv7 algorithm, four groups of ablation experiments are carried out as a way to verify the effectiveness of the proposed method, and the experimental results are shown in *Table 3*.

From *Table 3*, it can be concluded that after not utilizing the CBAM attention module, the detection precision rate, recall rate, and mean average precision are 74.1%, 63.8%,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136

and 65.5%, respectively; after adding only the CBAM attention module, the detection precision rate, recall rate, and mean average precision are 73.4%, 65.8%, and 66.5%, respectively. After using only the SIoU loss function, the detection precision rate, recall rate, and mean average precision are 77.9%, 62%, and 67.4%, respectively. When using both the CBAM attention module and the SIoU loss function, the detection precision rate, recall rate, and mean average precision are 78.8%, 65.4%, and 70.8%, respectively, which is 4.7%, 1.6%, and 5.3% higher than for the first set of data. It can be concluded that adding the CBAM attention module and introducing the SIoU loss function significantly improves the detection performance of this fire detection model. *Figure 7* shows the improved P-R curve. As shown in *Figure 8*, in 300 epochs training, the curve rapidly rises from 0 to 70 epochs and then tends to flatten, indicating that the model has basically converged and achieved optimal performance.

Table 3. *Ablation experiments*

Number	CBAM	SIoU	P	R	mAP@0.5
1	-	-	0.741	0.638	0.655
2	✓	-	0.734	0.658	0.665
3	-	✓	0.779	0.62	0.674
4	✓	✓	0.788	0.654	0.708

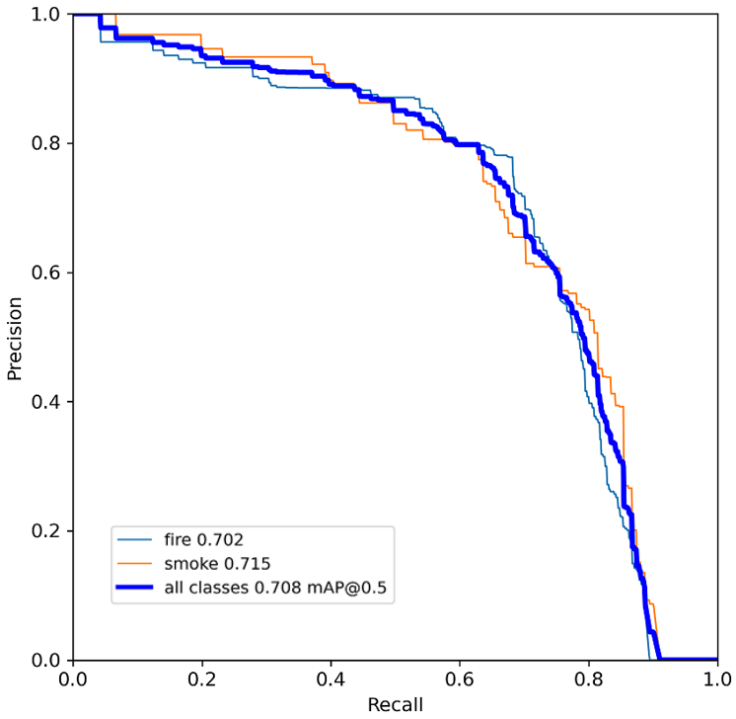

Figure 7. *Improved P-R curve.*

Comparative Tests

In order to demonstrate the superiority of the detection performance of the proposed method, it is tested in comparison with the current mainstream target detection algorithms YOLOv4 (Bochkovskiy et al., 2020) and YOLOv5 (Zhu et al., 2021). All the algorithms

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136

are trained and tested using the same training set and test set. The experimental results are shown in *Table 4* and *Figure 9*, Frames Per Second (FPS) is introduced to evaluate the detection speed of the model.

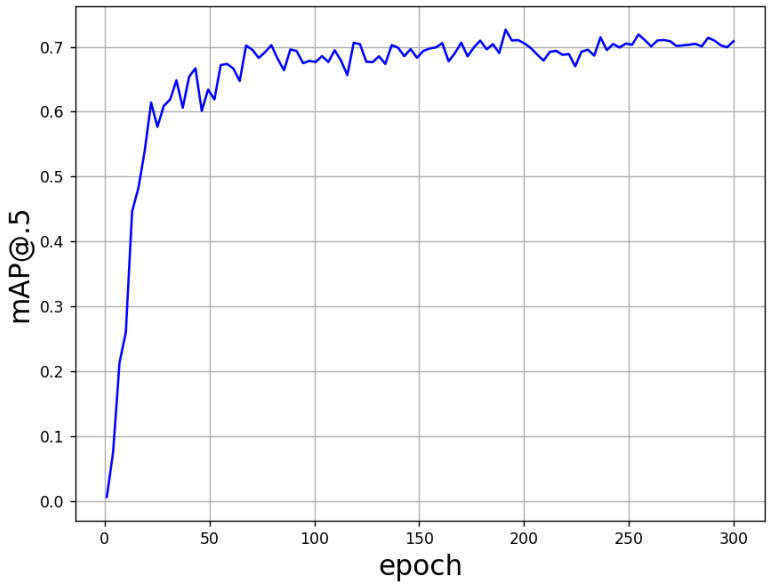

Figure 8. *Improved mAP@0.5-epoch curve*

Table 4. *Comparative experiments*

Algorithm	P	R	mAP@0.5	FPS
YOLOv4	0.28	0.50	0.36	78.53
YOLOv5	0.693	0.652	0.646	90.09
Ours	0.788	0.654	0.708	107.98

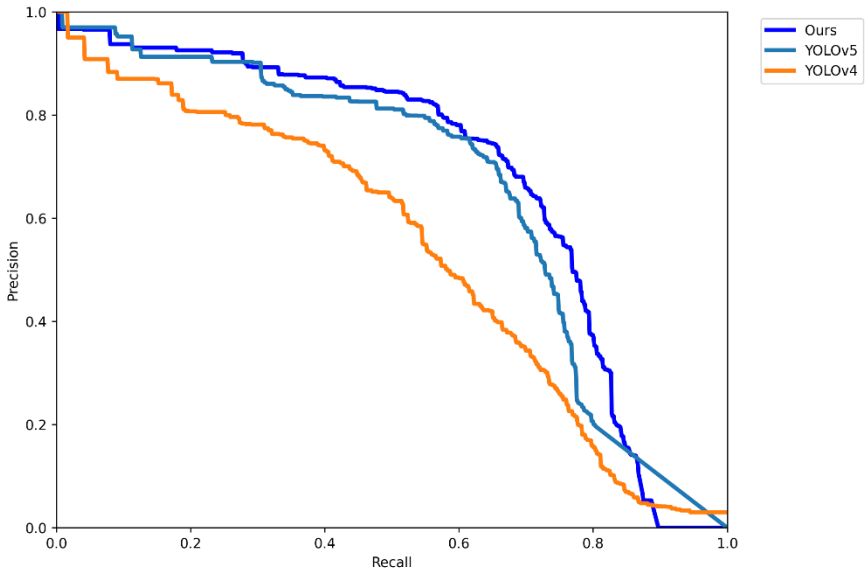

Figure 9. *P-R curves of comparative experiments*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136

From the comparison in *Table 4*, it can be seen that the improved algorithm is 50.8%, 15.4%, and 34.8% higher than YOLOv4 for the precision rate, recall rate, and mean average precision, respectively. Compared with YOLOv5, the precision rate, recall rate, and mean average precision for the proposed algorithm are 9.5%, 0.2%, and 6.2% higher, respectively. The FPS of our proposed method is 29.45 higher than YOLOv4 and 17.89 higher than YOLOv5. It is found that, in terms of the four detection indicators, the proposed algorithm is superior to the existing detection algorithms.

From the P-R curve in *Figure 9*, it can be seen that the dark blue curve corresponding to the improved algorithm in this paper is mostly higher than the light blue curve corresponding to YOLOv5, indicating that the overall performance of the model is better, and YOLOv4 is the worst.

After the improvement of the algorithm in this article, the model size is 11.7MB, and the size remains basically unchanged. Compared to other models, the algorithm speed has been significantly improved. According to the table, all four corresponding detection indicators have been improved.

Discussion

Firstly, the results demonstrate that the proposed improvements to the YOLOv7-tiny network effectively address the limitations of conventional fire detection techniques, which often suffer from delays, inaccuracies, and resource inefficiencies. By incorporating image enhancement, channel and spatial attention mechanisms, and the SIoU loss function, our model not only enhances the detection accuracy but also maintains real-time performance, thereby overcoming the timeliness issue associated with traditional approaches. The direct interaction with hardware, enabling instant response to environmental changes, further underscores the suitability of our model for real-time forest fire monitoring scenarios.

The introduction of image enhancement techniques significantly contributes to the model's ability to discern fire and smoke features in challenging forest environments, where lighting conditions, smoke obscuration, and complex backgrounds can hinder detection accuracy. The channel and spatial attention mechanisms, on the other hand, allow the model to selectively focus on salient features related to fire and smoke, enhancing its learning capacity, and leading to more precise detections. The employment of the SIoU loss function accelerates the bounding box regression, ensuring rapid localization of fire events, a crucial aspect for prompt intervention and mitigation.

The superior performance of our model in accurately identifying both smoke and fire signals a substantial advancement in forest fire management. Early smoke detection is particularly valuable, as it can serve as a precursor to a potential fire outbreak, enabling preventative actions before the situation escalates. This capability fills the gap in traditional methods that often neglect or struggle with smoke detection, thereby contributing to more comprehensive and effective fire protection measures.

While the results show promising outcomes, acknowledging the model's limitations and potential areas for further improvement is essential. Future research could explore the model's robustness under varying weather conditions, seasonal variations in vegetation, and different types of forest structures. Additionally, assessing the model's performance in detecting smoldering fires, which may produce less visible smoke, would provide a more comprehensive understanding of its applicability in diverse fire scenarios.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136

Integrating the model with multi-sensor data (e.g., thermal imaging, meteorological data) might further enhance its detection capabilities and overall reliability.

In terms of practical implementation, efforts should be directed towards optimizing model deployment strategies, ensuring seamless integration with existing monitoring systems, and addressing issues related to power consumption, maintenance, and connectivity in remote forested areas. Investigating the potential for edge computing solutions could reduce latency and minimize reliance on centralized processing infrastructure, enhancing the model's real-world usability.

Lastly, the societal and ecological implications of this work cannot be overstated. The successful deployment of our proposed model could lead to a marked reduction in property damage, loss of life, and environmental degradation caused by forest fires. By enabling timely detection and response, our model has the potential to contribute significantly to global efforts aimed at preserving forests and mitigating the impacts of climate change.

Conclusions

In order to improve the accuracy and real-time performance of forest fire recognition and detection, this paper proposes several improvement strategies for the YOLOv7-tiny network model. The algorithm first performs image enhancement for the imaging features, introduces the channel and spatial attention mechanism to improve the learning ability for the smoke and fire features in forest fires, and finally introduces the SIoU loss function to accelerate the bounding box regression of the model. The algorithm in this paper aims to improve the detection accuracy of the model while improving the model detection speed as much as possible. It has good engineering application capability as it can efficiently complete the high-precision identification of smoke and flames, realizing the prevention, control, and early warning for forest fires. In the follow-up work, the algorithm proposed in this paper will be investigated for model deployment and other aspects to further improve the practicality of the model, to meet the practical real-time requirements.

Acknowledgements. This work was supported by the Sichuan Science and Technology Program No. 2022YFG0325.

REFERENCES

[1] Abdusalomov, A. B., Islam, B. M. S., Nasimov, R., Mukhiddinov, M., Whangbo, T. K. (2023): An improved forest fire detection method based on the Detectron2 model and a deep learning approach. – Sensors 23(3): 1512.

[2] Alkhatib, A. A. A. (2014): A review on forest fire detection techniques. – International Journal of Distributed Sensor Networks 10(3).

[3] Bahhar, C., Ksibi, A., Ayadi, M., Jamjoom, M. M., Ullah, Z., Soufiene, B. O., Sakli, H. (2023): Wildfire and smoke detection using staged YOLO model and ensemble CNN. – Electronics 12(1): 228.

[4] Barmpoutis, P., Papaioannou, P., Dimitropoulos, K., Grammalidis, N. (2020): A review on early forest fire detection systems using optical remote sensing. – Sensors 20(22): 6442.

[5] Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y. M. (2020): YOLOv4: Optimal speed and accuracy of object detection. – ArXiv abs/2004.10934 (2020).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136

[6] Chen, S., Bao, H., Zeng, X., Yang, Y. (2003): A fire detecting method based on multi-sensor data fusion. – In: SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics 4: 3775-3780.
[7] Chen, T. H., Wu, P. H., Chiou, Y. C. (2004): An early fire-detection method based on image processing. – In: International Conference on Image Processing, 2004. ICIP'04 3: 1707-1710.
[8] Chen, G., Cheng, R., Lin, X., Jiao, W., Bai, D., Lin, H. (2023): LMDFS: A lightweight model for detecting forest fire smoke in UAV images based on YOLOv7. – Remote Sensing 15(15): 3790.
[9] Gevorgyan, Z. (2022): SIoU Loss: More Powerful Learning for Bounding Box Regression. – arXiv preprint arXiv: 2205.1 2740.
[10] Guede-Fernández, F., Martins, L., de Almeida, R. V., Gamboa, H., Vieira, P. (2021): A deep learning-based object identification system for forest fire detection. – Fire 4(4): 75.
[11] Jiao, Z., Zhang, Y., Xin, J., Mu, L., Yi, Y., Liu, H., Liu, D. (2019): A deep learning-based forest fire detection approach using UAV and YOLOv3. – In: 1st International Conference on Industrial Artificial Intelligence (IAI), Shenyang, China, pp. 1-5.
[12] Kim, S.-Y., Muminov, A. (2023): Forest fire smoke detection based on deep learning approaches and unmanned aerial vehicle images. – Sensors 23(12): 5702.
[13] Seydi, S. T., Saeidi, V., Kalantar, B., Ueda, N., Halin, A. A. (2022): Fire-Net: A deep learning framework for active forest fire detection. – Journal of Sensors, Article ID: 8044390.
[14] Wang, C. Y., Bochkovskiy, A., Liao, H. Y. M. (2022): YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. – Available at: arXiv: 2207.02696. [Accessed: 2022]. https://arxiv.org/abs/2207.02696.
[15] Woo, S., Park, J., Lee, J. Y., Kweon, I. S. (2018): CBAM: Convolutional block attention module. – In: Computer Vision - ECCV 2018, pp. 3-19. Cham: Springer International Publishing.
[16] Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D. W. (2020): Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. – Computer Science, Computer Vision and Pattern Recognition.
[17] Zhu, X., Lyu, S., Wang, X., Zhao, Q. (2021): TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios. – In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), pp. 2778-2788.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2204_31233136