Improved Model for Smoke Detection Based on Concentration Features using YOLOv7tiny

Yuanpan ZHENG[1]*, Liwei Niu[2], Xinxin GAN[3], Hui WANG[4], Boyang XU[5], Zhenyu WANG[6]
Zhengzhou University of Light Industry, Zhengzhou 450000, Henan, China[1,2,4,6]
SIPPR Engineering Group Co., Ltd, Zhengzhou 450007, Henan, China[3]
Zhengzhou University of Industrial Technology, Zhengzhou 451100, Henan, China[5]

***Abstract*—Smoke is often present in the early stages of a fire. Detecting low smoke concentration and small targets during these early stages can be challenging. This paper proposes an improved smoke detection algorithm that leverages the characteristics of smoke concentration using YOLOv7tiny. The improved algorithm consists of the following components: 1) utilizing the dark channel prior theory to extract smoke concentration characteristics and using the synthesized αRGB image as an input feature to enhance the features of sparse smoke; 2) designing a light-BiFPN multi-scale feature fusion structure to improve the detection performance of small target smoke; 3) using depth separable convolution to replace the original standard convolution and reduce the model parameter quantity. Experimental results on a self-made dataset show that the improved algorithm performs better in detecting sparse smoke and small target smoke, with mAP@0.5 and Recall reaching 94.03% and 95.62% respectively, and the detection FPS increasing to 118.78 frames/s. Moreover, the model parameter quantity decreases to 4.97M. The improved algorithm demonstrates superior performance in the detection of sparse and small smoke in the early stages of a fire.**

Keywords*—*YOLOv7tiny; smoke detection; dark channel; smoke concentration; feature fusion; depthwise separable convolution

I. Introduction

With the rapid development of the national economy and various industries, factories are producing more production materials, but they are also facing increased safety risks. High-density residential buildings are increasingly engaging in intensive fire and electricity usage behaviors. According to statistics from the Ministry of Emergency Management as of January 20, 2022, there were a total of 748,000 recorded fires in 2021, resulting in over 4,000 casualties and direct economic losses exceeding 6.75 billion yuan [1]. Therefore, it is crucial to research fire and smoke detection methods to ensure public property safety.

Currently, smoke detection research can be categorized into methods based on hardware sensors and wireless signals, and methods based on computer vision [2]. However, methods based on hardware sensors and wireless signals have poor adaptability in certain scenarios and do not perform well [3]. To overcome these limitations, computer vision-based smoke detection methods have been widely employed in recent years. Surveillance systems have also evolved from simulation-based, networked, and high-definition systems to intelligent systems. Now, surveillance resources are not only utilized for local monitoring functions but also integrated with computer vision for intelligent monitoring. Object detection algorithms based on deep learning have rapidly developed and become the mainstream method for smoke detection, as they possess powerful feature learning and representation capabilities, better meeting the requirements of the big data era in comparison to traditional machine learning methods [4].

He et al. [5] proposed a deep fusion convolutional neural network for smoke detection based on efficient attention, integrating spatial and channel attention mechanisms to address the issue of detecting small smoke. Sun et al. [6] presented an improved convolutional neural network for the rapid identification of forest fire smoke. However, the algorithm has poor generalization ability and weak robustness, only exhibiting high detection capability in specific scenarios. Wang et al. [7] proposed a smoke detection algorithm based on Faster R-CNN. Firstly, smoke is extracted based on its motion features, and then the Faster R-CNN network is used to extract and recognize the smoke image features, achieving high accuracy. However, the Faster R-CNN network structure is complex, and real-time detection is poor.

In recent years, the YOLO series models have garnered extensive research in the field of object detection due to their real-time performance, one-stage detection, simplicity, and good accuracy. Ren et al. [8] implemented fire detection and recognition using an improved YOLOv3 network. The algorithm improves the accuracy and detection speed of small smoke targets by modifying the predicted box sizes of the K-means clustering algorithm in YOLOv3. Cao et al. [9] proposed a precision enhancement strategy for YOLOv4 based on multi-scale feature maps and made improvements in detecting small objects by enhancing the feature extraction network. However, this significantly increased the algorithm complexity, resulting in a significant decrease in real-time detection. Xue et al. [10] proposed an improved model based on YOLOv5s. To address the issue of capturing effective information from small-sized targets in long-distance forest fire images, transfer learning methods were used to enhance the accuracy of small-target forest fire smoke detection. However, this model has a complex structure, and the detection accuracy is not sufficient [11].

The aforementioned fire smoke detection algorithms have improved the accuracy of smoke detection to some extent. However, they still face the following difficulties in the early stages of actual fire scenarios: 1) high false negative rate for thin smoke with a slow initial spread in fires; 2) difficulty in detecting small smoke targets captured from long distances; 3) high complexity of model algorithms, making real-time detection challenging.

To address these issues, this paper proposes a YOLOv7tiny lightweight improved network based on smoke concentration features, which significantly enhances the original network for complex smoke detection scenarios. The algorithm mainly includes 1) Extracting smoke concentration features based on the atmospheric transmission principle to enhance smoke characteristics and improve the detection capability for thin smoke; 2) Using a weighted bidirectional feature fusion structure to replace the original PAN+FPN feature fusion method, enhancing the algorithm's ability to detect small smoke targets; 3) Replacing the regular convolutions in the original network with depthwise separable convolutions with fewer parameters. The main contributions of this paper are:

1) Extracting smoke image concentration features based on the dark channel prior theory and enhancing the original RGB image to an αRGB image with smoke concentration features as the network input have been proven to enhance the detection capability of early-stage fires with thin smoke through experiments.

2) Proposing a lightweight feature fusion structure (light-BiFPN) to enhance the detection of small smoke targets in the YOLOv7tiny network and reduce the false negative rate of small smoke targets.

3) Replacing the standard convolutions of the original algorithm with depthwise separable convolutions, and experimental results show a significant reduction in parameters with minimal impact on accuracy.

Finally, the superiority of the proposed improvement algorithm was confirmed by analyzing the experimental results.

4) A dataset was created for detecting smoke objects in outdoor real-world scenes. The dataset comprises 1671 smoke images with corresponding labels indicating the position of the smoke bounding boxes. This dataset holds immense significance for researching the detection of smoke in the initial phases of actual fire scenarios.

II. Background

The YOLOv7 algorithm is a novel object detection algorithm introduced by the original development team of YOLOv4 in July 2022. Compared to previous versions of the YOLO series, this algorithm enhances the learning capability of the network through the use of the C5 module in the aggregation network. Additionally, it introduces attention mechanisms in the backbone feature extraction network to optimize the representation of target features, thereby achieving real-time detection. However, this algorithm has a relatively lower average precision. To achieve high-precision fire smoke detection in complex outdoor environments while reducing the number of algorithm parameters and improving detection speed, this study proposes improvements to the YOLOv7tiny algorithm. The improved algorithm includes the incorporation of a smoke concentration feature extraction structure and the use of a more lightweight multi-scale feature fusion network and optimized depthwise separable convolutions. With these enhancements, the algorithm can adapt to complex scenarios and achieve good real-time detection capability.

YOLOv7tiny is a deep learning-based object detection model composed of four parts: Input, Backbone, Neck, and Head. Fig. 1 shows the diagram of the YOLOv7tiny model. The Input part applies random mosaic data augmentation and K-means clustering to optimize the model training by designing anchor boxes for preprocessing the input images.

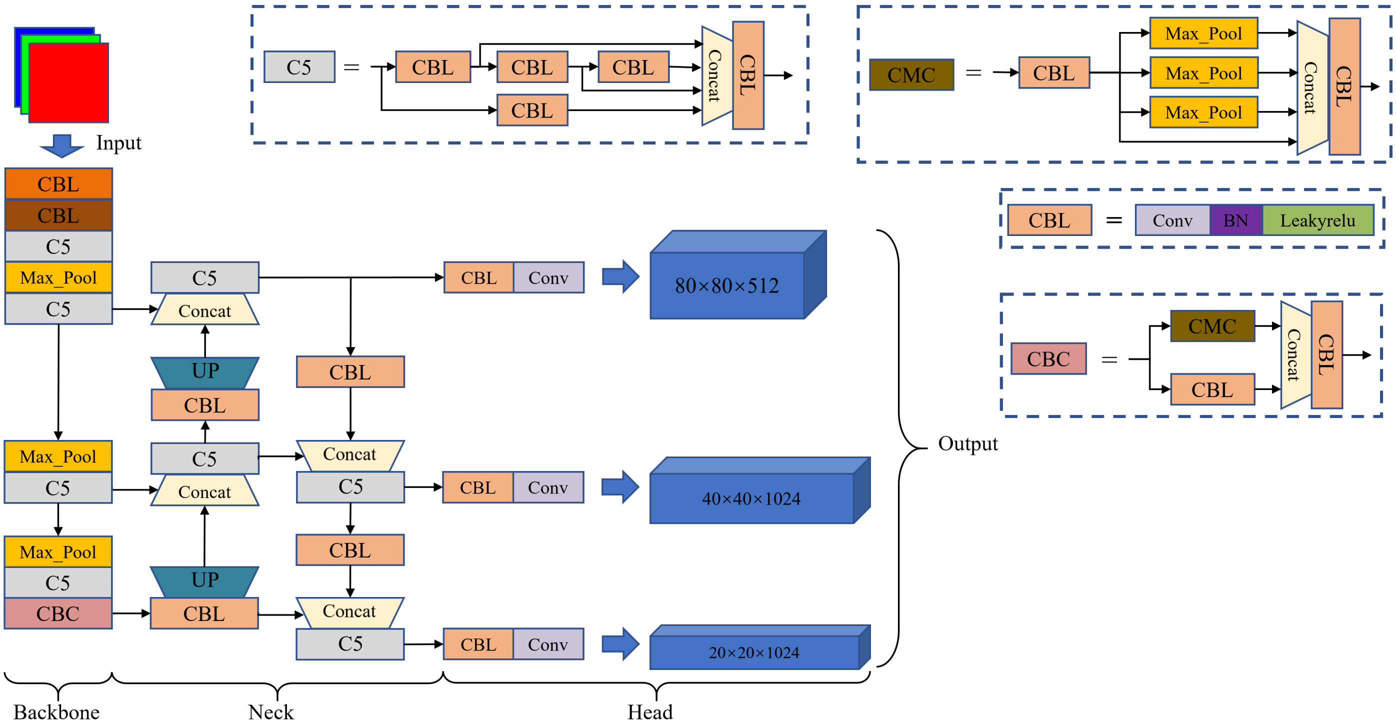


Fig. 1. The general architecture of the YOLOv7tiny network.

The Backbone part consists of multiple CBL modules, a C5 layer, and an MP layer. The CBL module is composed of a Convolution layer, a Batch Normalization layer, and a Leaky ReLU function. The C5 layer is formed by concatenating multiple CBL modules, and the MP layer includes CBL modules and Maxpool. The Neck part employs a feature fusion network, which adopts the YOLOv5 series Path Aggregation Feature Pyramid Network (PAFPN) architecture and combines Feature Pyramid Networks (FPN) [12] and Path Aggregation Networks (PAN) [13] to achieve multi-scale learning and retain small object features before downsampling. However, tensor concatenation for feature fusion lacks comprehensive integration of adjacent layer information, and nearest-neighbor interpolation for upsampling cannot effectively balance speed and accuracy in smoke detection tasks. The fusion network does not adequately focus on small object feature information, which can result in feature loss. The Head part uses a detection head similar to the YOLOR model, introducing the Implicit representation strategy [14] to refine the predictions. Based on the fused feature values, the images are classified into large, medium, and small categories, with the small image prediction branch primarily focusing on small defect objects. However, the detection head's use of IDetect to connect ordinary convolution prevents the fusion results from emphasizing the intended targets. Additionally, the detection head lacks targeted strategies to enhance small object detection performance.

III. Proposed Method

A. Smoke Concentration Feature Extraction Based on Dark Channel

Smoke concentration is a characteristic of smoke that directly reflects the content of smoke in the air per unit volume. In images, smoke concentration is closely related to the transmittance of the smoke image.

$$\alpha = 1 - t \tag{1}$$

Generally, the larger the smoke concentration (α), the smaller the transmittance of the image (t). The transmittance can be described by the smoke diffusion equation, which is a commonly used mathematical model for describing smoke concentration. Its form is as follows:

$$I = J \times t + A \times (1 - t) \tag{2}$$

Here, I represents the original foggy image, J represents the clear image after defogging, t represents the image transmission rate, and A represents the atmospheric light intensity. The dark channel prior theory [15] is a commonly used image defogging algorithm. It is based on the fog equation in Eq. (2) and analyzes the dark channel of the image to extract the transmission rate of the foggy image, thereby achieving image-defogging. The formula for the dark channel prior theory is as follows:

$$\min_{\Omega}(\min_{C}\frac{I^C}{A^C}) = \left\{\min_{\Omega}(\min_{C}\frac{J^C}{A^C})\right\} t + 1 - t \tag{3}$$

In the equation, I^C represents the RGB channels of the original foggy image, and J^C represents the clear and fog-free image. Through the analysis conducted by He et al. [15], it has been revealed that the majority of images in real outdoor fog-free scenes have a significant amount of dark channels with very low pixel values, i.e., $\min_{\Omega}(\min_{C}\frac{J^C}{A^C}) \to 0$. Therefore, after simplifying Equation (3), we can proceed with the processing:

$$t = 1 - \min_{\Omega}(\min_{C}\frac{I^C}{A^C}) \tag{4}$$

In the equation, Ω represents the sliding window size. First, the brightest region is searched in the dark channel image, and then the brightness of the corresponding region in the original image is taken as the atmospheric light intensity (A^C). As a result, the transmission rate of the foggy image (t) can be calculated.

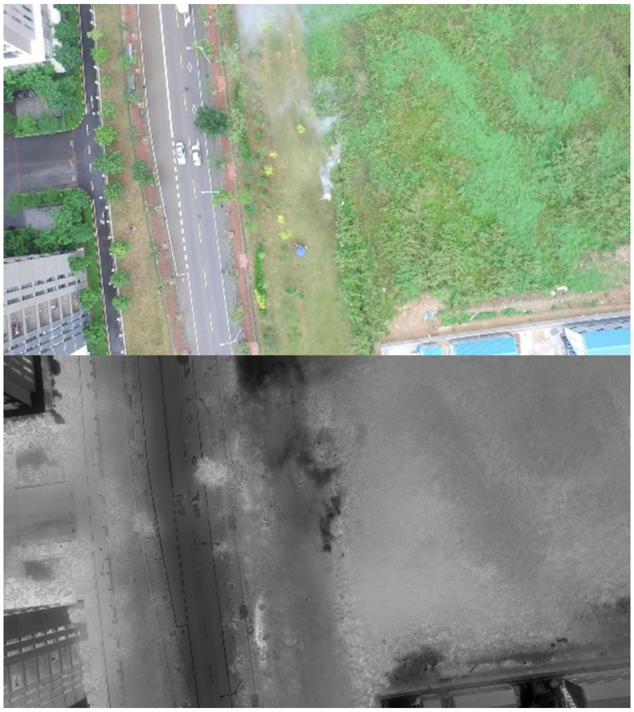


Fig. 2. Smoke image and corresponding transmittance grayscale image.

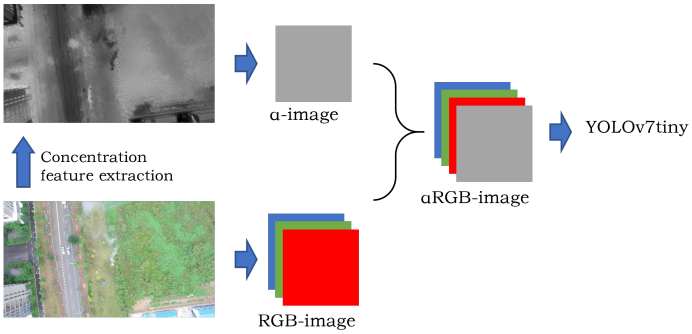


Fig. 3. Extraction of concentration features.

Mo et al. [16] extracted smoke transmittance based on the smoke aerosol equation and measured smoke concentration under different lighting conditions. The experiments demonstrated that using smoke transmittance for estimating smoke concentration is feasible and accurate. According to Eq. (4), the transmittance of smoke images can be calculated pixel by pixel. Mapping the transmittance of smoke to a grayscale image allows for a visual representation of the transmittance map as shown in Fig. 2 (bottom).

The transmittance grayscale image exhibits dark areas that indicate low transmittance, suggesting a blockage of light, similar to smoke particles. Consequently, the smoke transmittance image is combined with the RGB image of the smoke, creating a four-dimensional vector as the input for the network model. This merged αRGB image, as depicted in Fig. 3, retains the original image's shape, color, and texture, while also reflecting the inherent concentration features of the smoke.

B. Improved Feature Fusion Network

Fig. 4 shows the PAN, BiFPN, and light-BiFPN feature fusion structures. In object detection tasks, feature fusion plays a crucial role in enhancing model accuracy. Traditional feature fusion methods focus on top-down and bottom-up feature propagation processes, with the PAN structure (Fig. 4(a)) being the most representative method [13]. By cascading, the PAN structure merges feature information from different levels and scales to expand the model's receptive field and improve detection accuracy. Its main advantage lies in effectively leveraging information from features of various scales to obtain a richer and more accurate representation.

However, the PAN structure does have some deficiencies when dealing with small objects, which can be manifested in the following two aspects:

1) **Feature Information Loss**: When merging feature information from different levels and scales, the PAN structure is prone to information loss, especially impacting the detection performance of small objects.
2) **Unstable Fusion Effects**: The cascading approach utilized in the PAN structure tends to encounter problems like gradient vanishing or explosion, leading to unstable feature fusion effects.

To address these issues, this study replaces the original PAN structure with the light-BiFPN (Bidirectional Feature Pyramid Network) [17]. The BiFPN structure [Fig. 4(b)] introduces lateral connections during the top-down and bottom-up fusion processes, effectively enhancing the exchange and transmission of feature information, particularly improving the detection performance of small targets.

The BiFPN structure is composed of multiple cascaded BiFPN modules. Each module comprises two feature propagation paths (top-down and bottom-up) and lateral connection paths. During the feature propagation process, the BiFPN module adopts a multi-level feature fusion approach to combine features from multiple sizes. These fused features are then passed to the subsequent module until the final module outputs the ultimate feature map. The lateral connection paths employ learnable weights to facilitate effective feature fusion between different layers. The weights of lateral connections are obtained through convolutional operations. Assuming the input feature map is x_i, the weights of lateral connections w_{ij} can be denoted as:

$$w_{ij} = ReLU(W_{ij}[x_i, x_j]) \tag{5}$$

Here, W_{ij} represents a learnable weight matrix, and $[x_i, x_j]$ signifies the concatenation of feature maps x_i and x_j.

Compared to the PAN structure, the BiFPN structure better preserves detailed information of small targets, thereby improving the accuracy and robustness of object detection. Additionally, due to the scalability of the BiFPN structure, different numbers of modules and structural parameters can be chosen according to the specific scenario to achieve optimal detection performance. The light-BiFPN used in this study [Fig. 4(c)] reduces the feature layers P6 and P7 to reduce model parameters and optimize model speed.

C. Depthwise Separable Convolution

In object detection algorithms, convolutional neural networks (CNN) are commonly employed as backbone networks to extract image features. Standard convolution serves as

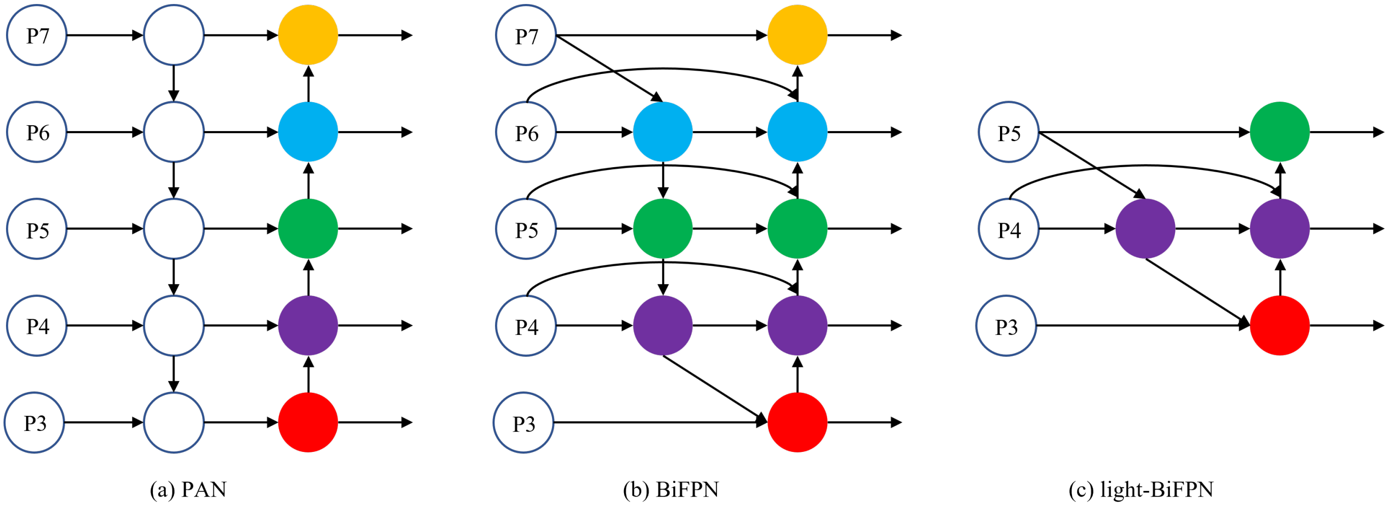


Fig. 4. PAN, BiFPN, and light-BiFPN feature fusion structures.

one of the most prevalent CNN modules, extracting features by conducting convolution operations on input feature maps and convolution kernels. However, when dealing with large-scale feature maps, standard convolution leads to significantly increased computational and memory consumption, restricting the depth and complexity of the model.

To address this issue, this paper adopts Depthwise Separable Convolution (DSC) [18] as a replacement for standard convolution in the backbone network. Depthwise Separable Convolution decomposes the standard convolution into depthwise convolution and pointwise convolution, performing convolution operations on each channel and each pixel of the input feature map, respectively. This approach considerably reduces the number of parameters and computations while ensuring the accuracy and efficiency of the model.

Depthwise Separable Convolution can be represented by the following formula:

$$Y\& = PW(DW(X)) \tag{6}$$

Where X represents the input feature map, DW represents the depth convolution operation, PW represents the pointwise convolution operation, and W represents the parameters of the convolutional kernel. The depth convolution operation and the pointwise convolution operation correspond to two independent convolutional layers, with parameter quantities of D_k and $D_k \times D_o$, respectively. Here, D_k represents the number of channels in the input feature map, K represents the size of the convolutional kernel, and D_o represents the number of channels in the output feature map. Compared to standard convolution, depthwise separable convolution reduces the parameter and computational requirements by K^2 and D_k times, respectively.

IV. Experimental and Result Analysis

To test the effectiveness of the improved algorithm, training and testing were conducted on a self-made dataset. The optimization effects of various improvements were analyzed through horizontal comparative experiments and vertical ablation experiments.

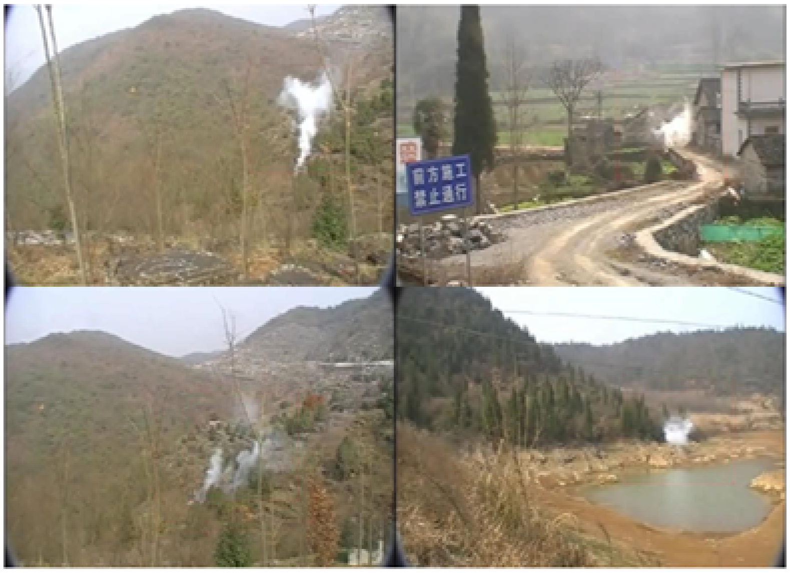
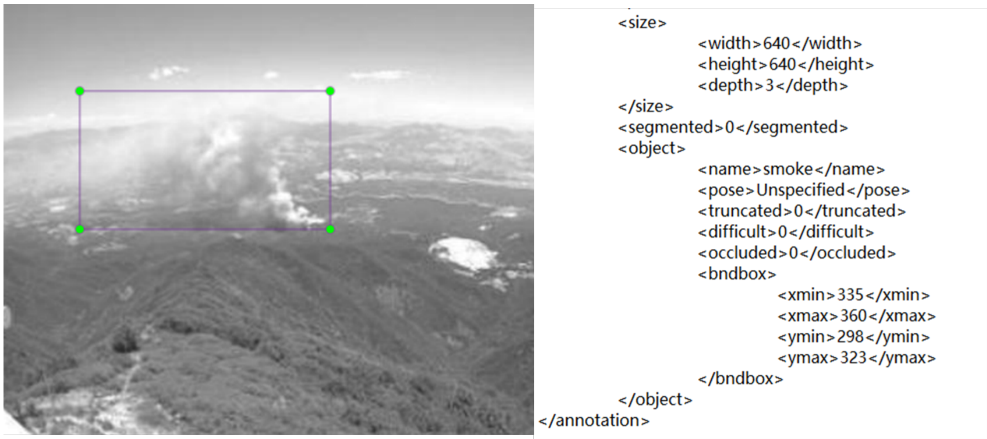


Fig. 5. Sample images from a self-made fire smoke detection dataset.



```
<size>
        <width>640</width>
        <height>640</height>
        <depth>3</depth>
</size>
<segmented>0</segmented>
<object>
        <name>smoke</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <occluded>0</occluded>
        <bndbox>
                <xmin>335</xmin>
                <xmax>360</xmax>
                <ymin>298</ymin>
                <ymax>323</ymax>
        </bndbox>
</object>
</annotation>
```

Fig. 6. Example of image annotation.

A. Dataset and Preprocessing

Currently, there is a limited availability of publicly accessible outdoor real fire smoke datasets. In this study, 3604 unlabeled smoke images were collected from publicly available smoke image datasets, as shown in Fig. 5. After removing low-quality images, 1671 smoke images were selected and manually annotated using the LabelImg tool to create a self-made smoke detection dataset in Pascal VOC2007 format. Fig. 6 demonstrates the smoke targets and their corresponding XML information. The dataset was split as follows:

$$(TrainingSet + ValidationSet) : TestSet = 9 : 1$$

$$TrainingSet : ValidationSet = 9 : 1$$

The training set, validation set, and test set consist of 1352, 151, and 168 images, respectively.

In the data preprocessing stage of this experiment, in addition to using traditional image processing techniques such as image flipping and HSV color space enhancement, random mosaic, and mixup image processing techniques [19] were also applied to randomly augment the dataset, aiming to enhance the robustness of the model.

The random mosaic technique combines multiple images into a new image to enhance the diversity of the dataset, while the mixup technique linearly blends two different images to generate a new image. Both data augmentation techniques effectively increase the sample size of the dataset, improving the model's generalization ability and further enhancing the accuracy of smoke object detection.

B. Experimental Environment and Parameter Settings

1) Hardware and software Environment

The experimental hardware environment of this article is shown in Table I.

TABLE I. EXPERIMENTAL ENVIRONMENT

CPU	AMD EPYC 7773X @ 3.50GHz
GPU	GeForce RTX 3090
RAM	30G
Operating System	Ubuntu
Programming Language	Python 3.8
Deep Learning Framework	PyTorch 1.8
GPU Acceleration Library	CUDA 11.1

2) Training Hyperparameters Settings

The experimental hyperparameter settings of this article are shown in Table II.

TABLE II. TRAINING HYPERPARAMETERS SETTINGS

Hyperparameter	Value
Mosaic Probability	0.5
Mixup Probability	0.5
Maximum Learning Rate	0.01
Minimum Learning Rate	0.0001
Epoch	300

During the training process, the VOC pre-trained weights of YOLOv7tiny were utilized. The first 50 epochs comprised of frozen training, where only the Neck and Head parts' parameters were trained while the backbone feature extraction network remained frozen. From epoch 51 to 300, the unfrozen training stage occurred, and the entire network was trained. The batch size was set to 64 during the frozen training stage, and it was reduced to 32 during the unfrozen training stage to accommodate the increase in training parameters. The cosine learning rate decay method was employed to progressively decrease the learning rate from 0.01 to 0.0001. The stochastic gradient descent method with a momentum of 0.937 was chosen as the parameter optimizer. Additionally, a weight decay coefficient of 5e-4 was implemented to prevent overfitting during the training process.

C. Evaluation Metrics

To evaluate the performance of the improved algorithm, this study uses four metrics for algorithm assessment: Recall, mean Average Precision (mAP), Frames Per Second (FPS), and model parameter count (Params).

1) Recall: Recall measures the detection rate of a model for all true positive samples. In smoke object detection tasks, the calculation formula for Recall is as follows:

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

Where TP represents true positive, referring to the number of positive samples correctly detected by the model, while FN represents false negative, indicating the number of positive samples that the model fails to detect. Recall is utilized in this paper as one of the evaluation metrics to assess the detection capability of the algorithm.

2) mAP: mAP stands for mean Average Precision, which measures the average precision of a model at different confidence thresholds. In smoke object detection tasks, the formula to calculate mAP is as follows:

$$mAP = \frac{1}{n}\sum\nolimits_{i=1}^{n} AP_i \tag{8}$$

Where n represents the number of classes, and AP_i represents the average precision of the ith class. In this paper, since only smoke is involved as the target, mAP can be considered as AP, used to evaluate the detection accuracy of the algorithm.

3) FPS: FPS stands for Frames Per Second, which measures the number of frames processed by a model per unit of time. In the smoke detection task, the calculation formula for FPS is as follows:

$$FPS = \frac{1}{t} \tag{9}$$

In this case, t represents the average time for processing a frame image. This paper utilizes Frames Per Second (FPS) as one of the evaluation metrics to assess the detection speed of the algorithm.

4) Params: The model parameter count refers to the number of trainable parameters in the model, which is an important indicator for evaluating model complexity. In the task of smoke object detection, the calculation formula for model parameter count is given by Eq. 10.

$$N = \sum\nolimits_{i=1}^{n} (w_i h_i c_i k_i^2 + b_i) \tag{10}$$

TABLE III. Ablation Experiment Results

Experimental Number	Improvement			Evaluation Metric			
	αRGB	Light-BiFPN	DSC	Recall	mAP@0.5	FPS	Params(M)
1	✖	✖	✖	88.21%	89.48%	106.65	6.23
2	✔	✖	✖	91.12%	92.33%	95.40	6.23
3	✖	✔	✖	91.99%	91.54%	94.46	6.31
4	✖	✖	✔	86.63%	87.80%	**124.68**	**4.82**
5	✔	✔	✔	**95.62%**	**94.03%**	118.78	4.97

Here, n represents the number of layers in the model. w_i, h_i, and c_i represent the width, height, and number of channels of layer i, respectively. k_i represents the size of the convolutional kernel in layer i, and b_i represents the bias term in layer i. In this study, the model complexity is evaluated based on the number of model parameters.

D. Ablation Experiment

To validate the benefits of each improvement point on the network model, five ablation experiments were conducted. The experimental environment and parameter settings were kept consistent. The results of the ablation experiments are shown in Table III.

1) The first set of experiments is conducted using the YOLOv7tiny algorithm, serving as a comparative benchmark for the subsequent improvement experiments.

2) The second group of experiments is a control experiment with the inclusion of smoke concentration features. By introducing smoke concentration features, the computational burden of the model increases, resulting in a decrease in the detection frame rate. However, it achieved good performance in terms of Recall and mAP, with improvements of 2.91 and 2.85 percentage points, respectively.

3) By analyzing the experimental data of the first and third groups, it is concluded that the light-BiFPN structure increases the number of model parameters due to the addition of skip connections, which leads to a decrease in the detection frame rate. However, it demonstrates good performance in terms of accuracy and recall rate, with improvements of 3.78 and 2.06 percentage points, respectively.

4) The fourth set of experiments replaced the standard convolution in the original YOLOv7tiny network model with depthwise separable convolution (DSC). Analyzing the experimental data compared to the baseline network reveals that DSC can significantly reduce the number of parameters and improve the detection frame rate. However, the reduced number of parameters limits the expressive power of the model.

5) In the fifth experiment, the improved YOLOv7tiny network based on smoke concentration features proposed in this paper is evaluated. From the experimental data, it can be observed that compared to the baseline network, the Recall and mAP have improved by 7.41 and 4.55 percentage points, respectively. The FPS has improved by 12.13 frames/s, and the number of parameters has decreased from 6.23M to 4.97M. Therefore, it can be concluded that the algorithm proposed in this paper is lighter and more accurate.

E. Comparative Experiment

To investigate the performance of the improved network in detecting different targets, this study conducted three sets of comparative experiments on a self-made dataset: comprehensive comparison experiment, smoke concentration comparison experiment, and multi-scale target comparison experiment. To comprehensively assess the performance of the algorithm, mainstream object detection models were selected as the comparison models, including RetinaNet [20], CenterNet [21], EfficientDet [22], Faster R-CNN [23], SSD [24], and YOLOv5s [25].

1) Comprehensive comparative experiment: A comprehensive comparative experiment was conducted by training six mainstream detection algorithms on a self-made dataset for 300 epochs as comparison algorithms to the proposed algorithm in this paper. From the variations of mAP@0.5 of each algorithm during the training process (shown in Fig. 7), it can be observed that, apart from the proposed algorithm, Faster R-CNN and YOLOv5s performed remarkably well on this dataset. Both the proposed algorithm and Faster R-CNN converged quickly (basically converged at 50 epochs). The proposed algorithm achieved an mAP of 94.03% after final convergence, surpassing other algorithms.

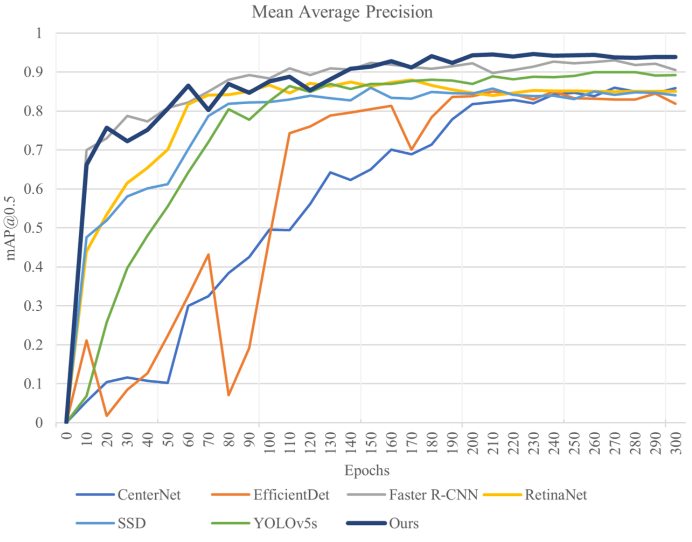


Fig. 7. mAP@0.5 Variation Graph of Each Algorithm during Training Process.

2) Smoke Concentration Comparison Experiment: The concentration features are extracted from the smoke images in the dataset. Then, the mean concentration of the smoke region can be obtained by calculating the average of the concentrations within the smoke bounding box. The distribution of smoke concentrations in this dataset is shown in Fig. 8.

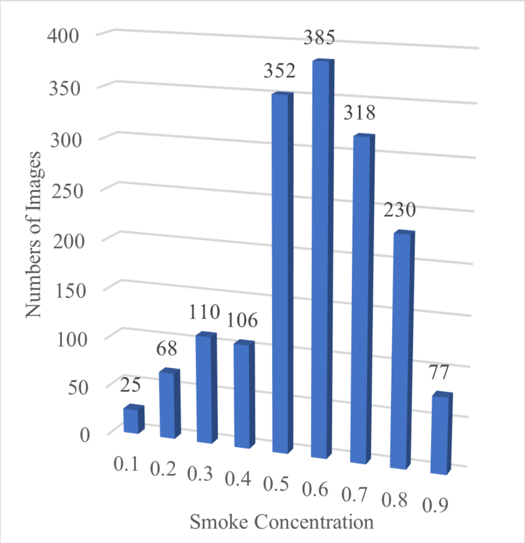


Fig. 8. Histogram of Smoke Target Concentration Distribution.

In Fig. 8, the x-axis represents the smoke concentration, and the y-axis represents the number of smoke images corresponding to each concentration. In this experiment, the smoke concentration is divided into low concentration and high concentration. The low concentration is defined as below 0.5, and the high concentration is defined as 0.5 and above. The performance of the improved model and mainstream object detection models are compared on the low-concentration and high-concentration test image sets.

TABLE IV. EXPERIMENTAL RESULTS OF SMOKE CONCENTRATION COMPARISON

Compare Models	Low Concentration mAP(%)	High Concentration mAP(%)	Params (M)
RetinaNet	84.75%	88.01%	36.33
CenterNet	83.35%	87.64%	32.67
EfficientDet	82.56%	86.51%	**3.83**
Faster R-CNN	90.45%	93.67%	136.69
SSD	82.96%	87.26%	23.61
YOLOv5s	86.75%	90.43%	46.63
Ours	**93.27%**	**94.64%**	4.97

From Table IV, it can be observed that both the mainstream algorithm models and our proposed improved algorithm model achieve similar detection accuracy for high-concentration smoke. However, our improved algorithm model achieves a significant reduction in parameter size, down to 4.97M. This reduction is particularly important for deploying the model on edge devices. Ordinary algorithms struggle to distinguish low-concentration smoke due to its semi-transparent nature. In contrast, our improved algorithm achieves good performance on low-concentration smoke, thanks to the introduced αRGB concentration feature.

3) Multi-scale Object Comparison Experiment: In the early stage of a fire, the smoke volume is usually small. However, the detection of smoke in the early stage is particularly important for firefighting. Therefore, a small object comparison experiment is designed to test the performance of different algorithms in detecting smoke from small objects.

Using the K-means algorithm, a cluster analysis of the size of smoke targets in the dataset was performed. The average silhouette coefficient was found to be 1.74, and the center points corresponded to large, medium, and small targets with sizes of 33×23, 80×60, and 160×142, respectively. Fig. 9 shows the distribution of width and height for the three scales of smoke targets in the self-made dataset. The horizontal axis represents the width of the smoke target, and the vertical axis represents the height of the smoke target.

The scale analysis of 168 smoke images in the test set reveals that there are 36 large objects, 47 medium-sized objects, and 85 small objects. As shown in Fig. 10, it can be observed that small smoke objects occupy a significant portion. The detection results of various algorithm models on this test set are shown in Fig. 10.

From the multi-scale object comparison experimental results in Fig. 11, it can be seen that although CenterNet, EfficientDet, SSD, and YOLOv5s have higher mAP in detecting large and medium objects, they are slightly inferior in detecting small objects. RetinaNet and Faster R-CNN perform well in detecting objects of different scales, but overall, the mAP is relatively low. By using improved algorithms, especially the optimization of light-BiFPN, the detection accuracy of small objects is significantly improved, and they have higher mAP in object detection at various scales.

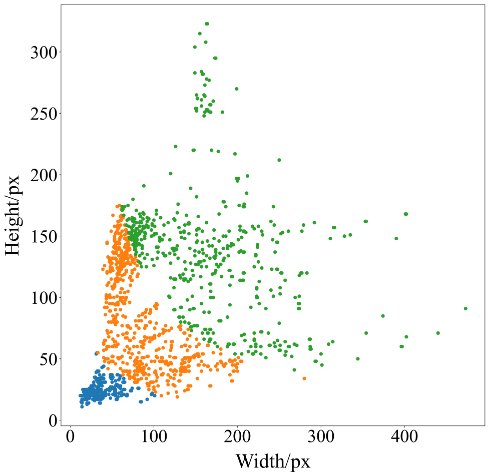


Fig. 9. Distribution of smoke objects in self-made dataset.

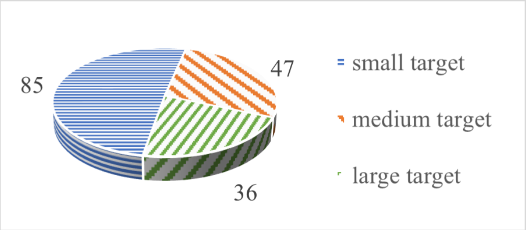


Fig. 10. Proportion of smoke objects at various scales in the test set.

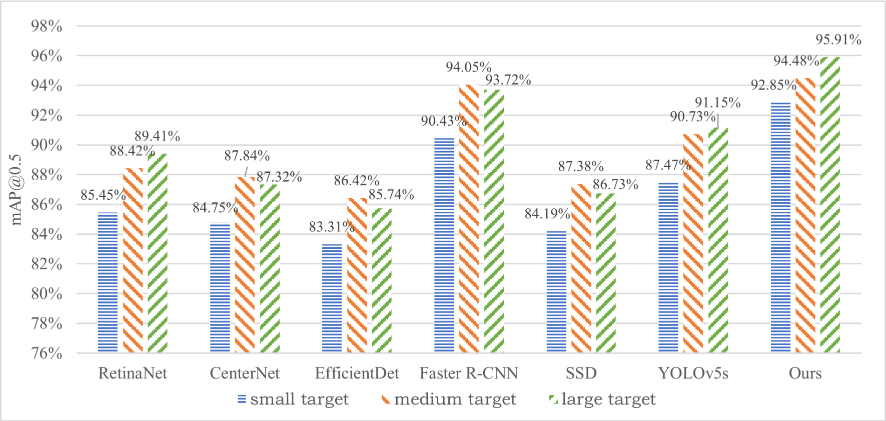


Fig. 11. Multi-scale object comparison experimental results.

Based on the multi-scale object comparison experimental results shown in Fig. 11, it can be observed that CenterNet, EfficientDet, SSD, and YOLOv5s have higher mean Average Precision (mAP) in detecting large and medium objects. However, they are slightly inferior in detecting small objects. RetinaNet and Faster R-CNN perform well in detecting objects of different scales but have relatively low overall mAP. On the other hand, our improved algorithms, especially with the optimization of light-BiFPN, achieve significantly better detection accuracy for small objects and higher mAP in object detection at various scales.

F. Detection Performance Analysis

The YOLOv7tiny algorithm performs poorly in detecting sparse smoke due to its low concentration in the early stages of a fire. This is because sparse smoke appears semi-transparent, often leading to false alarms [Fig. 12(a), Fig. 12(b)], missed detections [Fig. 12(d)], and low detection accuracy [Fig. 12(c)]. However, after introducing the αRGB feature, our algorithm significantly improves the detection capability of sparse smoke [Fig. 13(a-d)]. Nonetheless, due to the limited proportion of low-concentration smoke in the dataset, occasional cases may arise where the detected bounding boxes do not align with the actual ones [Fig. 13(f)]. By optimizing the light-BiFPN, our algorithm achieves more accurate detection of small targets [Fig. 13(d)] and performs closer to ideal in complex environments [Fig. 13(e)].

Fig. 12 shows the detection performance of the YOLOv7tiny algorithm, while Fig. 13 depicts the detection performance of our improved algorithm.

V. Conclusion

Fire and smoke detection plays a significant role in ensuring fire safety. By combining computer vision technology to accurately locate early-stage smoke in a fire, it serves as an important tool for fire warning and prevention of fire spread. To improve the detection of small and sparse smoke in the early stages of a fire, this study extracts features related to smoke concentration, improves feature fusion structures, and optimizes algorithm complexity. By comparing with other models on a self-made dataset, the recall rate reaches 95.62%, mAP reaches 94.03%, and the detection FPS is increased to 118.78. The algorithm complexity is reduced to 4.97M. The experimental results demonstrate the superiority of the improved algorithm in detecting sparse smoke. In future work, the algorithm will be further optimized in two aspects: firstly, by increasing the proportion of sparse smoke in the dataset to enhance algorithm robustness; secondly, by attempting to enhance algorithm expression capability through attention mechanisms, thereby further improving detection accuracy.

Acknowledgment

This work was partially supported by the Science and Technology Research Projects of Henan Province (No.222102210021), and the Key Scientific Research Project Plan for Higher Education Institutions of Henan Province, China (No.23A520004).

References

[1] “In 2021, the number of fire incidents reached a record high, with 745,000 fire extinguishments.”National Fire and Rescue Administration, Jan. 2022. https://www.119.gov.cn/gk/sjtj/2022/26442.shtml.

[2] J. He, L. Li, H. Lin, and G. Xu, “Overview of Research on Smoking Detection Methods in Computer Vision.”Computer Engineering and Applications, pp. 1-19. 2023.

[3] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, “A Forest Fire Detection System Based on Ensemble Learning.”Forests, vol. 12, no. 2, pp. 1-17, Feb. 2021.

[4] A. Yazdi, H. Qin, C. Jordan, L. Yang, and F. Yan, “Nemo: An Open-Source Transformer-Supercharged Benchmark for Fine-Grained Wildfire Smoke Detection.”Remote Sensing, vol. 14, no. 16, pp. 3979, Aug. 2022.

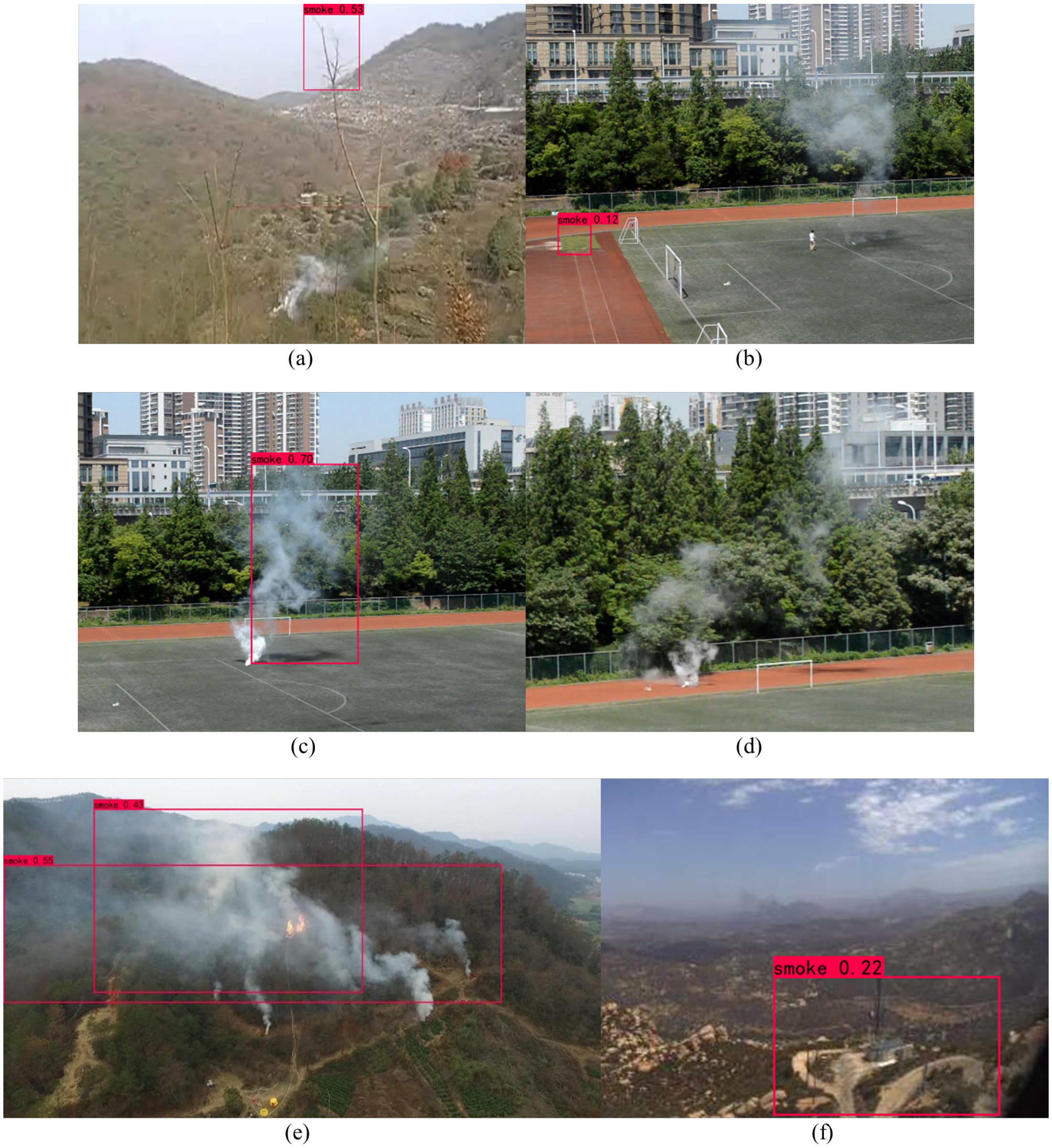
[5] L. He, X. Gong, S. Zhang, L. Wang, and F. Li, “Efficient Attention Based Deep Fusion CNN For Smoke Detection in Fog Environment.”Neurocomputing, vol. 434, pp. 224-238, Apr. 2021.

[6] X. Sun, L. Sun, and Y. Huang, “Forest Fire Smoke Recognition Based on Convolutional Neural Network.”Journal of Forestry Research, vol. 32, no. 5, pp. 1921-1927, Oct. 2021.

[7] F. Wang, “Research and Implementation of Forest Fire Detection System Based on Deep Learning.”University of Electronic Science and Technology of China, 2022

[8] J. Ren, W. Xiong, Z. Wu, and M. Jiang, “Fire detection and identification based on improved YOLOv3.”Computer System and Application, vol. 28, no. 12, pp. 171-176, Dec. 2019.

[9] C. Cao, X. Tan, X. Huang, Y. Zhang, and Z. Luo, “Study of flame detection based on improved YOLOv4.”Journal of Physics: Conference Series, vol. 1952, no. 2, Jun. 2021.



(a) (b)

(c) (d)

(e) (f)

Fig. 12. YOLOv7tiny Detection Performance.

[10] Z. Xue, H. Lin, and F. Wang, “A small target forest fire detection model based on YOLOv5 improvement.”Forests, vol. 13, no. 8, pp. 1332, Aug. 2022.

[11] A. Sukumaran, and T. Brindha,“Nature-inspired hybrid deep learning for race detection by face shape features.”International Journal of Intelligent Computing and Cybernetics, vol. 13, no. 3, pp. 365-388, Aug. 2020.

[12] T. Lin, P. Dollár, R. GIRSHICK, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection.”IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936-944, 2017.

[13] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path aggregation network for instance segmentation.”IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8759-8768, 2018.

[14] C. Wang, I. Yeh, and H. LIAO, “You Only Learn One Representation: Unified Network for Multiple Tasks.”arXiv preprint arXiv:2105.04206 (2021).

[15] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior.”IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no, 12, pp. 2341-2353, Dec. 2011.

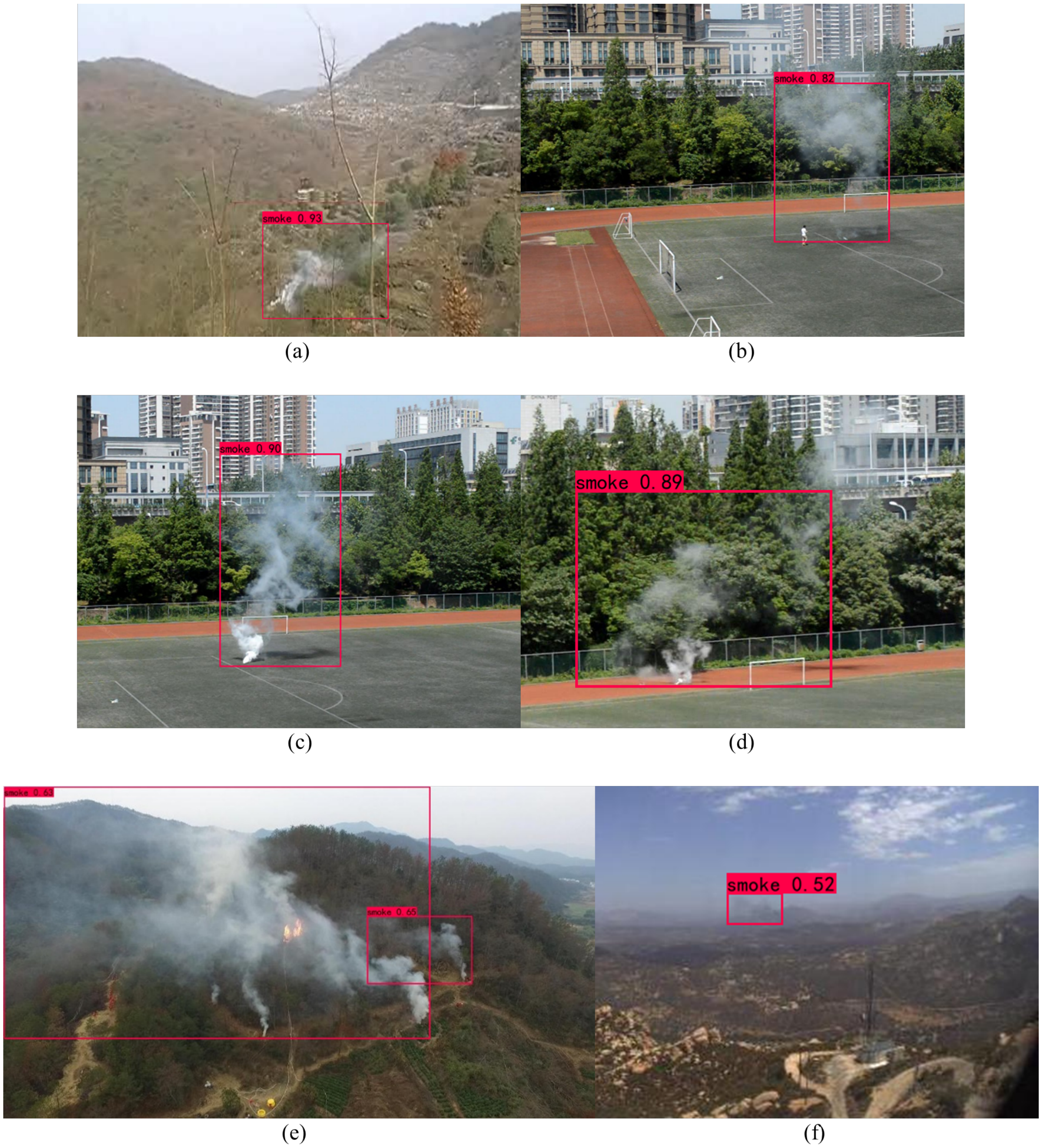
[16] H. Mo, and Z. Xie, “Smoke Concentration Measurement Method Based on Dual-Channel Deep CNN.”Pattern Recognition and Artificial Intelligence, vol. 34, no. 9, pp. 844-852, Sep. 2021.

[17] M. Tan, R. Pang, and Q. Le, “EfficientDet: Scalable and Efficient Object Detection.”IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.10778-10787, 2020.

[18] A. Howard, M. Sandler, B. Chen, W. Wang, L. Chen et al., “Searching for Mobilenetv3”IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1314-1324, 2019.

[19] H. Zhang, M. Cisse, Y. Dauphin, and D. Lopez-Paz, “MixUp: Beyond empirical risk minimization.”6th International Conference on Learning Representations (ICLR), 2018.

[20] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal Loss for Dense Object Detection.”IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 2, pp. 318-327, Feb. 2020.



(a) (b)

(c) (d)

(e) (f)

Fig. 13. Improved YOLOv7tiny detection performance.

[21] X. Zhou, J. Zhuo and P. Krähenbühl, “Bottom-Up Object Detection by Grouping Extreme and Center Points.”IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 850-859, 2019.

[22] M. Tan, R. Pang and Q. V. Le, “EfficientDet: Scalable and Efficient Object Detection.”IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10778-10787, 2020.

[23] S. Ren, K. He, R. Girshick and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.”IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 39, no. 6, pp. 1137-1149, Jun. 2017.

[24] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg, “SSD: Single shot multibox detector.”Lecture Notes in Computer Science, vol. 9905, pp. 21-37, 2016.

[25] “yolov5.”Ultralytics, 2021. https://github.com/ultralytics/yolov5.