Detecting Forest Fires by UAVs Based on Improved YOLOv7

Yicheng Jin

School of Information Science and Engineering, East China University of Science and Technology, shanghai, China
22013348@mail.ecust.edu.cn

Abstract. The purpose of this study is to construct an improved forest fire detection model to improve the accuracy and real-time performance of UAV fire detection in forest environment. By introducing MobileNetv3 into YOLOv7 as the backbone network, its lightweight and efficient feature extraction capabilities can be used to optimize the model to lower its computational overhead and improve the detection speed, so as to ensure that the UAV can quickly respond to fire signals with limited hardware resources. The innovation of this research lies in breaking through the limitations of the traditional YOLOv7 backbone network architecture, using MobileNetv3 to realize the lightweight of the model, optimizing the feature fusion and connection mode, strengthening the model's learning capability for forest fire features, reducing the computational complexity by 61% and improving the average accuracy and speed, more accurately identifying fire signs in complex forest environments, and improving the overall efficiency of the UAV fire detection system.

Keywords: YOLOv7, MobileNetv3, forest fire.

1 Introduction

1.1 Background and Significance

Against the backdrop of increasing global ecological concerns, the protection of forest resources has become a key component of sustainable development. As a major threat to forest ecosystems, the detection, prevention, and control of forest fires have received much attention from countries. According to relevant statistics, the frequency of forest fires has been on the rise in recent years, causing serious damage to ecological balance, biodiversity, and economic development. According to the data related to forest fires released by the National Bureau of Statistics for the three years from 2020-2022, the situation of forest fires in China is relatively severe. The highest number of forest fires was 1,153, and the affected forest area was as much as 6,853.88 hectares, and there were even 44 casualties due to the disaster. Through Table 1, it can be clearly understood that forest fires can greatly threaten and damage forest resources, ecosystems, and human life and property safety. Therefore, it is especially critical and urgent to actively carry out and implement forest fire prevention work [1].

J. Iqbal (ed.), *Proceedings of the 2025 2nd International Conference on Mechanics, Electronics Engineering and Automation (ICMEEA 2025)*, Advances in Engineering Research 272,
https://doi.org/10.2991/978-94-6463-821-9_65

UAVs have emerged as a crucial tool for wildfire detection due to their agility and extensive coverage. For instance, Australia employs Lockheed Martin Space Systems Company to survey fires and terrain, aiding in firefighting efforts [2]. Similarly, the United States has implemented UAV programs through agencies like the Forest Service to enhance fire suppression mobility and response efficiency. Developed nations are integrating UAV development into their strategic plans, continuously improving their capabilities. However, traditional methods, such as satellite remote sensing and ground-based monitoring stations, suffer from detection blind spots and delayed responsiveness. Current technologies face challenges in accurately identifying early-stage smoke and small flames, as well as in optimizing model lightweighting.

Table 1. National Forest Fire Situation from 2020 to 2022.

Time (Year)	Forest fire incidents (times)	Affected forest area (hectares)	Casualties in forest fires (people)
2022	709	6853.88	44
2021	616	4456.62	28
2020	1153	8526.2	41

1.2 Research on YOLO algorithm in the field of UAV fire detection

In recent years, deep learning technology has experienced rapid development and has been widely applied in the field of forest fire detection. Compared to traditional forest fire detection methods, deep learning methods do not require significant human resources and demonstrate significant advantages in terms of high adaptability and high accuracy, playing an increasingly important role in the practice of forest fire detection. The YOLO algorithm performs well in object detection within the domain of deep learning. Its efficient architectural design can effectively extract and integrate features, reducing computational complexity and processing time, which is advantageous for fire early warning and rescue efforts, outperforming traditional methods and some other models. However, it still has limitations, such as the inability to adapt to specific scene requirements, and poor detection results when facing complex fire scenarios and small target detection problems. Therefore, improving the algorithm by introducing new technologies and structures has become a new research direction.

For example, Norkobil et al. [3] proposed a YOLOv8-based network architecture, using Wise-IoUv3 as the bounding box regression loss and replacing the traditional convolutional process of the intermediate neck layers with a Ghost Shuffle Convolution mechanism, achieving an average precision of 79.4% in detecting forest fire smoke. Wang et al. [4] introduced a UAV-based forest fire image detection method utilizing YOLO-TF, with CSPDarknet as the backbone network. Empirical evaluations on the FireDet-2025 dataset validated that the proposed framework outperformed the YOLOv5 baseline. Zhao et al. [5] proposed a UAV forest fire recognition method based on saliency detection and deep learning. This method processes the original images using an algorithm that combines saliency detection with a logistic regression classifier to locate the core fire area and segment the images. The model achieved 98% accuracy on the validation set and exhibited superior accuracy and generalization capabilities compared to other methods and in tests using actual news report images, outperforming some existing approaches. Luan et al. [6] presented a method for small object wildfire detection in UAV images based on an improved YOLOX. Through the integration of multi-level CSP-ML feature extraction architecture, the embedding of CBAM attention mechanism in the neck network, and the introduction of adaptive feature fusion module, a statistically significant improvement in object detection performance was achieved—a 6.4% improvement in the 0.5 intersection union ratio (mAP@0.5) and a 2.17% improvement in the average accuracy (mAP@0.5:0.95) within the 0.5-0.95 IoU threshold range.

This study introduces a novel structural enhancement algorithm, presenting an optimized YOLOv7-based approach for forest fire detection. By leveraging lightweight and efficient feature extraction capabilities, the paper aim to reduce model computational complexity and enhance detection speed, thereby enabling rapid response to fire signals from UAVs operating under resource constraints.

2 Forest Fire Detection Model Based on Combination of Mobilenetv3 and Yolov7

2.1 Introduction to the YOLOv7 algorithm

The YOLO (You Only Look Once) algorithm, introduced by Redmon Joseph et al. at CVPR 2016, formulates object detection as a regression task, mapping image pixels directly to bounding box coordinates and class probabilities. This single-stage detector, employing a unified convolutional neural network, enables end-to-end optimization of detection performance [7]. Subsequent iterations, including YOLOv5, YOLOv6, and YOLOv7, have been developed. YOLOv5 incorporates lightweight network architectures, achieving enhanced speed and accuracy. YOLOv6 is specifically optimized for industrial applications [8, 9]. Furthermore, Wang et al. [10] proposed YOLOv7 at CVPR 2023, which introduces innovative techniques such as re-parameterized label assignment, the ELAN efficient network architecture, and auxiliary head training. YOLOv7 maintains its speed advantage while improving detection accuracy through advancements in the backbone network and feature fusion methods.

2.2 Introduction to MobileNetv3

In 2019, Google introduced MobileNetV3, a lightweight convolutional neural network architecture [11]. The primary innovation lies in the application of Neural Architecture Search (NAS). This was achieved by integrating platform-aware NAS with NetAdapt. Specifically, platform-aware NAS, also known as block-wise search, optimizes network modules under constraints on model parameters and computational complexity. NetAdapt, or layer-wise search, fine-tunes each network layer after module determination, particularly adjusting the number of convolutional kernels per layer. Following model determination via NAS and NetAdapt, the computational cost of the initial and final layers was found to be high. To enhance detection speed while maintaining accuracy, network structure optimization was required. MobileNetV3 employs depthwise separable convolutions to construct a lightweight neural network, making it suitable for hardware devices and mobile terminals. This network integrates the depthwise separable convolution concept from MobileNetV1, the inverted residual structure with linear bottlenecks from MobileNetV2, and additionally incorporates the Squeeze-and-Excitation (SE) module, resulting in an improved detection network model.

2.3 Development of a Forest Fire Detection Model Based on the MobileNetv3-YOLOv7 Algorithm

The paper integrated the lightweight MobileNetv3 architecture into the YOLOv7 framework. Modifications were made to the input and output connections to ensure the proper integration of MobileNetv3's feature maps into the YOLOv7 neck [12]. The resulting parameter statistics and GFLOPS comparisons are presented in Table 2.

Based on the results in Table 2, it can be concluded that the number of model parameters using the MobileNetv3 network is smaller than that of YOLOv7, which means that the MobileNetv3-YOLOv7 model is able to perform better with a limited amount of data and may be faster to train. While YOLOv7 is able to learn more complex function mappings, it also requires more data for training, otherwise it is prone to overfitting. Also comparing the GFLOPS, it can be seen that compared to the YOLOv7 model, the MobileNetv3-YOLOv7 model needs to perform fewer floating-point operations per second when performing forward propagation and backpropagation calculations, which means that it is less computationally complex and requires less hardware computational power.

Table 2. Comparison of Model Parameter Statistics and GFLOPS.

Network model	Parameter statistics (10,000)	GFLOPS (floating point operations per second)
YOLOv7	3720.19	105.1
MobileNetv3-YOLOv7	2546.61	41.1

3 Result

3.1 Training Preparation

The experimental setup utilized a CUDA 12.4 environment with an RTX 4090 GPU (24GB). The dataset, sourced from online repositories, comprised a total of 4,000 images. This dataset was partitioned into training, validation, and test sets, with an 8:1:1 allocation ratio, respectively. Additional key training parameters are detailed in Table 3.

Table 3. Network hyperparameters.

Parameters	Numerical values
epochs	600
batch size	16
img size	416
initial learning rate	0.01

Data preprocessing was performed, involving image rotation, brightness adjustment, and contrast modification. Figures 1 and 2 present a subset of the training dataset, designed to simulate the unstable image acquisition conditions typical of UAV-based aerial photography. Furthermore, synthetic smoke noise was manually introduced.

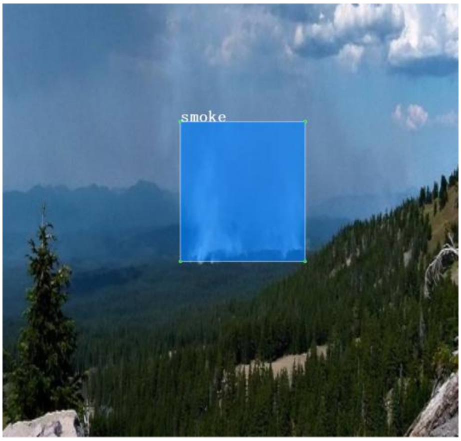

Fig. 1. Original image without data augmentation.

Fig. 2. Enhanced image after data augmentation.

3.2 Test Results

Following the completion of the training phase, the model's performance was evaluated using the test dataset. Test images were fed into the trained model to generate object detection results. These results were then compared to assess the model's performance metrics.

The terms True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are fundamental for representing classification outcomes, comparing model predictions against ground truth labels. Specifically, TP denotes the count of fire-related pixels correctly identified as positive, aligning with positive ground truth labels. TN represents the number of non-fire pixels accurately classified as negative, consistent with negative ground truth. FP indicates the number of pixels classified as fire, but are, in reality, negative according to the ground truth. FN stands for the count of pixels that are determined to be non - fire through the classification process. Despite being labeled as positive in the ground truth. Intersection over Union (IoU) is also a critical evaluation metric, quantifying the overlap between two bounding boxes. In object detection tasks, models predict bounding boxes for target objects, and IoU assesses the agreement between these predicted boxes and the ground truth annotations.

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{1}$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{2}$$

$$\mathrm{AP} = \int \mathrm{P(R)dR} \tag{3}$$

$$\mathrm{mAP} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{AP_i} \tag{4}$$

This study employs Precision and Recall as primary evaluation metrics. Furthermore, the paper compares the performance of two models based on their mean Average Precision (mAP) at an IoU threshold of 0.5, and across an IoU range from 0.5 to 0.95. The results are presented in Tables 4 and 5.

Table 4. The evaluation results of YOLOv7.

YOLOv7	P	R	mAP@.5	mAP@.5:.95
Smoke	0.76	0.63	0.59	0.36
Smoke+Fire	0.70	0.62	0.59	0.31

Table 5. The evaluation results of MobileNetv3-YOLOv7.

MobileNetv3-YOLOv7	P	R	mAP@.5	mAP@.5:.95
Smoke	0.76	0.65	0.68	0.43
Smoke+Fire	0.69	0.63	0.64	0.36

3.3 Data Analysis

The Precision-Recall (P/R) curve is a critical tool in information retrieval and machine learning for evaluating the performance of binary classification models. The curve is typically generated by plotting recall on the x-axis and precision on the y-axis, traversing all possible classification thresholds. A larger area under the curve indicates superior model performance. Figures 3 and 4 show the P/R curves of both models respectively and the forest fire image detection results on the test dataset are presented in Figure 5.

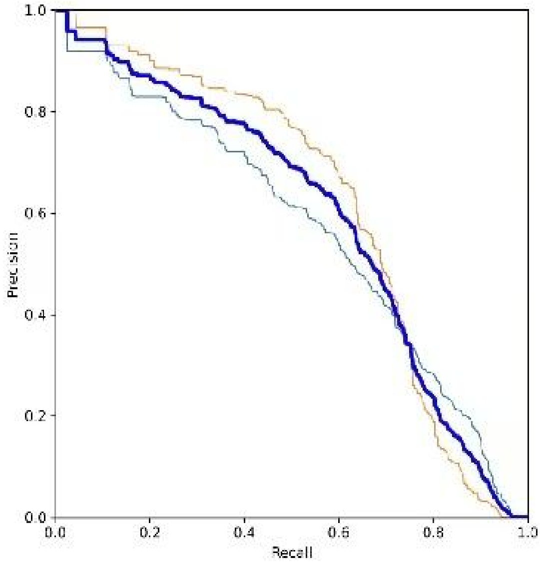

Fig. 3. MobileNetv3-YOLOv7 P/R curve.

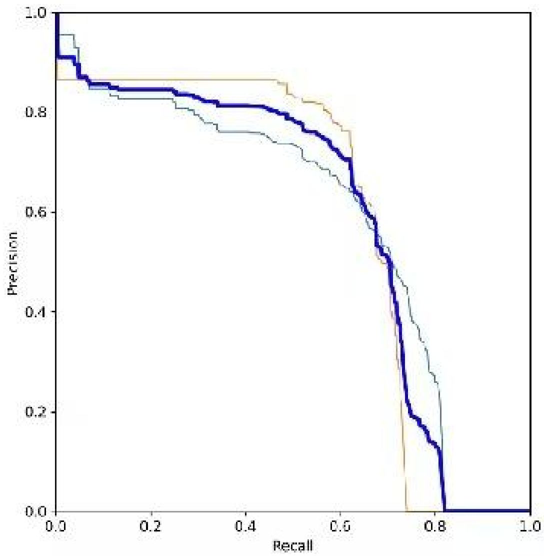

Fig. 4. YOLOv7 P/R curve.

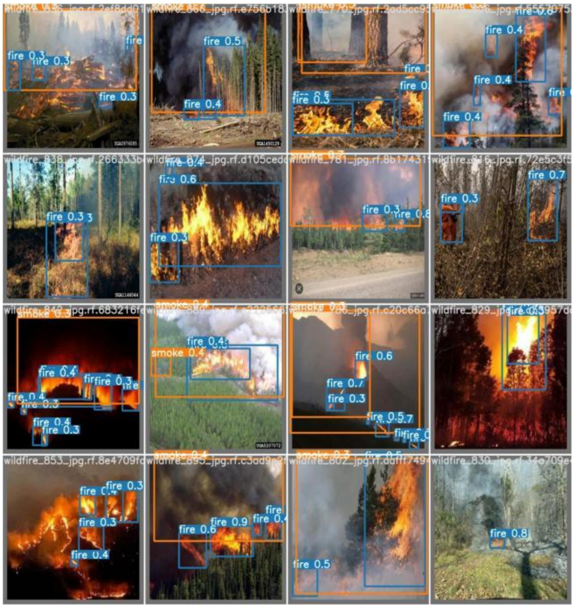

Fig. 5. MobileNetv3-YOLOv7 Test set image classification results display.

Comparative analysis reveals that MobileNetv3-YOLOv7 algorithm exhibits a marginal decrease in precision (P) but a slight increase in recall (R) compared to the baseline. Examination of the precision-recall (PR) curves in Figures 3 and 4 indicates that the MobileNetv3-YOLOv7 curve is positioned closer to the upper-right quadrant overall. This suggests a superior balance between precision and recall across varying recall levels. Conversely, YOLOv7 demonstrates a more pronounced decline in precision at both low and high recall scenarios. Consequently, MobileNetv3-YOLOv7 outperforms YOLOv7 in terms of the integrated performance of precision and recall. Analysis of the mean Average Precision (mAP) results, specifically mAP@.5 and mAP@.5:.95, presented in Tables 4 and 5, demonstrates that MobileNetv3-YOLOv7 offers an advantage in the precise localization and delineation of smoke targets. However, both models exhibit potential for improvement in high-precision detection of fire+smoke targets. Overall, MobileNetv3-YOLOv7 achieves enhanced recall and mAP metrics at the expense of a 1% reduction in precision. Furthermore, the lightweight design results in a 61% reduction in computational complexity, thereby significantly improving detection speed. Therefore, the proposed MobileNetv3-YOLOv7 algorithm is more suitable for practical forest fire detection deployment applications compared to the conventional YOLOv7 algorithm [13].

4 Conclusion

In this study, the paper put forward a modified YOLOv7 algorithm by integrating MobileNetv3 network as the backbone feature extraction module. This modification simplifies the network architecture, thereby reducing the parameter count and computational load. This approach offers a novel technical perspective and methodology for forest fire detection using unmanned aerial vehicles (UAVs). The paper constructed forest fire detection models based on both the original and modified algorithms and validated their performance using a forest fire image dataset. The results indicate that our proposed algorithm, while sacrificing 1% in precision, significantly enhances recall and mAP metrics while concurrently reducing computational complexity. This leads to a substantial improvement in detection speed, thereby better meeting the real-time requirements of forest fire detection.

Despite the significant improvements in detection accuracy and speed, the model's performance exhibits some variability under extremely complex environmental conditions, such as dense smoke and strong light interference. Furthermore, the current model's ability to recognize different types of fire characteristics requires enhancement, particularly for specific fire types like ground fires and crown fires, where the model's adaptability is limited. Future investigations should aim to achieve quantifiable improvements to improve its robustness and generalization capabilities in complex environments, enabling it to better handle various fire scenarios. Integrating multi-source data (e.g., meteorological data, terrain data) and multi-modal information (e.g., images, videos, sensor data) could facilitate the construction of a more comprehensive and accurate forest fire detection model.

References

1. Zhentian, J.: Research on UAV system for forest fire prevention', Master thesis, Xi'an University of Technology (2019).
2. Shan, B., Mingfei, S., Binhai, Z., et al.: The development tendency of UAV for forest fire protection at home and abroad, Forestry and Grassland Machinery, **29**(03):34-36 (2018).
3. Saydirasulovich, S.N., Mukhiddinov, M., Djuraev, O., et al: An improved wildfire smoke detection based on yolov8 and uav images, Sensors (Basel), **23**(20):8374 (2023).
4. Anding, W.: Research on UAV forest fire image detection algorithm based on deep learning, Civil Aviation Flight University of China (2023).
5. Zhao, Y., Ma, J., Li, X., et al.: Saliency Detection and Deep Learning-Based Wildfire Identification in UAV Imagery'. Sensors (Basel). **18**(3):712 (2018).
6. Luan, T., Zhou, S., Zhang, G., et al.: Enhanced Lightweight YOLOX for Small Object Wildfire Detection in UAV Imagery, Sensors (Basel). **24**(9): 2710 (2024).
7. Redmon, J., Divvala, S., Girshick, R., et al.: You Only Look Once: Unified, Real-Time Object Detection', Computer Vision & Pattern Recognition. IEEE (2016).
8. Shuo, W., Yangyang, L., Xinhua, W., et al.: An improved YOLO algorithm for UAV detection in formation flight, Nanjing University of Aeronautics and Astronautics (2025).
9. Chuyi, L., Lulu, L., Hongliang, J., et al.: YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications, Shanghai AI Laboratory (2022).
10. Wang, Y., Bochkovskiy, A., Liao, H.Y M.: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, arXiv e-prints (2022).
11. Howard, A., Sandler, M., Chen B., et al.: Searching for MobileNetV3, 2019 IEEE/CVF International Conference on Computer Vision (ICCV) IEEE (2020).

12. Danni, T.: Research on deep learning methods for forest fire detection, Xi'an University of Technology (2021).
13. Mojtaba, N., Mohsen, A., Hossein, E.: Vehicle Type and Speed Detection on Android Devices Using YOLO V5 and MobileNet, Islamic Azad University (2024).