A Lightweight Forest Fire Detection Method Based on UAV Dual-Modal Images

Lingxia Mu, *Member, IEEE*, Yichi Yang, Youmin Zhang, *Fellow, IEEE*, Xianghong Xue, and Nan Feng

***Abstract*—This letter presents a lightweight method for detecting forest fires using dual-modal remote sensing images captured by an uncrewed aerial vehicle (UAV). The aim is to achieve efficient fire monitoring on a computationally resource-constrained UAV platform. The proposed detection network is based on the improved YOLOv8, which uses RGB image and thermal image as network input at the same time. A lightweight dual-modal feature fusion module named dual-modal fusion module (DFM) is designed to effectively combine RGB and thermal features. The existing C2f module in YOLOv8 was replaced by the lightweight module C2f-F, along with the addition of the parameter-free attention module SimAM. This improvement improves the detection performance of the model while minimizing the model parameters. The evaluation experimental results on the FLAME 2 dataset show that the accuracy of the proposed dual-modal forest fire detection method reaches 98.4%, and the model size is only 2.9 MB, which achieves a good balance between accuracy and number of parameters compared with other mainstream methods. In addition, on the iCrest 2-s edge computing device, the detection speed reaches 20.67 frames per second (FPS), further confirming that this lightweight approach satisfies the real-time detection requirements for forest fires.**

***Index Terms*—Edge computing device, forest fire detection, improved YOLOv8, lightweight network, remote sensing, RGB and thermal image, uncrewed aerial vehicles (UAVs).**

I. Introduction

FORESTS are a fundamental component of terrestrial ecosystems, effectively storing carbon while supporting rich biodiversity. Moreover, they provide society with ecosystem products and services worth hundreds of billions of dollars each year [1]. Therefore, the protection of forest resources can not only provide renewable resources for human activities but also plays an important role in ecological protection and global energy cycle [2]. Unfortunately, vast areas of forest are damaged by fires every year. These incidents pose a significant natural threat, threatening ecosystems, economic attributes, infrastructure, and human life [3]. Given the quick spread and prolonged burning associated with forest fires, early detection is recognized as one of the most effective strategies for mitigating losses caused by these disasters [4].

Forest fire monitoring technology can usually be divided into ground monitoring, satellite remote sensing monitoring, and airborne remote sensing monitoring. Ground monitoring usually has a fixed monitoring area and can be affected by the forest topography. With the development of satellite remote sensing technology, it is possible to use remote sensing images to monitor ground targets. In [5], a precise real-time detection algorithm was designed to detect coconut trees using satellite images. However, while satellite remote sensing monitoring can quickly collect data over large areas, there may be delays in capturing early fire information. With the rapid development of uncrewed aerial vehicle (UAV) technology, airborne remote sensing monitoring has attracted extensive attention [6]. UAVs enable more effective fire monitoring due to their flexibility and wide detection range [7].

With advancements in sensor technology, UAV equipped with optics and other sensors are able to provide high-quality fire images or videos. Traditional image-based detection methods usually include three steps: image preprocessing, feature extraction, and classification [8]. These methods depend on manually designed features, which limits their adaptability to various environments and hinders real-time performance, making it challenging to effectively respond to evolving fire situations. In contrast, forest fire detection methods based on deep learning can identify fire characteristics in complex images more accurately through automatic feature extraction and network training, which provides a more effective solution for forest fire monitoring [9].

Deep-learning-based detection methods typically use information provided by RGB images or thermal images for fire detection. In [10], a network model based on YOLOv5 was designed for the UAV platform for fire detection. In [11], a YOLO-based lightweight fire detection network is designed for use on embedded devices. Considering that temperature is another key feature of fire, [12] designed a semantic segmentation network based on mask R-CNN and YOLO to detect forest fires through thermal images. In [13], a CNN-based forest fire detection method is proposed, which uses images captured by near-infrared cameras for fire detection.

To take advantage of both RGB image and thermal image, a dual-modal image fire detection network based on ResNet50

Received 28 May 2025; accepted 14 June 2025. Date of publication 17 June 2025; date of current version 24 June 2025. This work was supported in part by the Aeronautical Science Foundation of China under Grant 2024Z0710T6002 and in part by the Open Fund of Key Laboratory of the Ministry of Education on Artificial Intelligence in Equipment under Grant 2024-AAIE-KF03-02. *(Corresponding author: Xianghong Xue.)*

Lingxia Mu, Yichi Yang, and Xianghong Xue are with Shaanxi Key Laboratory of Complex System Control and Intelligent Information Processing, Xi'an University of Technology, Xi'an, Shaanxi 710048, China (e-mail: mulingxia@xaut.edu.cn; xhxue@xaut.edu.cn).

Youmin Zhang is with the Department of Mechanical, Industrial and Aerospace Engineering, Concordia University, Montreal, QC H3G 1M8, Canada (e-mail: ymzhang@encs.concordia.ca).

Nan Feng is with the School of Intelligence Science and Technology, University of Science and Technology Beijing, Beijing 100083, China (e-mail: fengnan@ustb.edu.cn).

Digital Object Identifier 10.1109/LGRS.2025.3580564

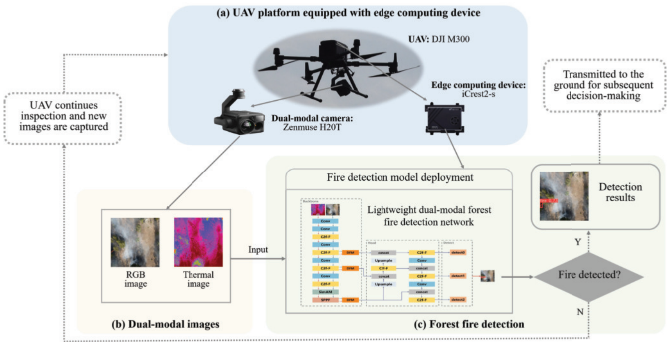

Fig. 1. Structure of the UAV-based dual-modal image forest fire detection.

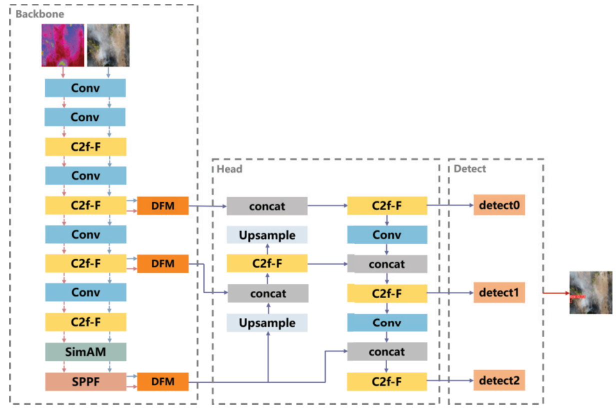

Fig. 2. Structure of the lightweight dual-modal forest fire detection network.

was proposed in [14]. By designing a dual-modal feature fusion module, RGB and thermal features are fully integrated to achieve accurate fire detection. However, the previously mentioned dual-modal fire detection methods often overlook the importance of network lightweighting, which poses challenges for real-time deployment of UAV platforms. Therefore, a lightweight dual-modal fire detection method is designed, which achieves high detection accuracy and uses fewer network parameters. The main contributions of this letter are as follows.

1) Aiming at real-time forest fire detection tasks, a novel real-time forest fire detection method based on UAV dual-modal remote sensing images was proposed. The response efficiency to forest fires monitoring can be achieved.
2) To enhance detection accuracy in complex environments, a fire detection algorithm that integrates RGB and thermal images has been developed. A lightweight dual-modal feature fusion module was proposed for edge devices with limited computing resources to achieve efficient dual-modal image fusion, thereby improving fire detection accuracy.
3) To better deploy the detection model on the edge computing device mounted on the UAV platform, a lightweight convolution module and a parameter-free attention module are integrated into the proposed dual-modal fire detection framework. This combination minimizes the number of model parameters while ensuring high detection accuracy.

The proposed UAV-based fire detection method is described in Section II. Section III describes the results of experiments performed on publicly available datasets and their analysis. Finally, the main contributions of this article are summarized in Section IV.

II. UAV-Based Forest Fire Detection Method

Fig. 1 shows the scheme of the UAV-based forest fire detection method. Among them, the solid arrows represent the flow of data between hardware components such as drones, dual-modal cameras, and edge devices. The dashed arrows indicate conditional actions, and if a fire is detected, the results are sent to the ground control center, otherwise monitoring continues. The blue part represents the drone platform and sensor, the orange part represents the dual-modal image, and the green part represents the algorithm deployment on the edge computing device. The method uses the UAV platform to carry edge computing equipment and combines the proposed lightweight dual-modal forest fire detection network to achieve efficient and real-time forest fire monitoring and early warning. As shown in Fig. 1(a), the DJI M300 drone equipped with a Zenmuse H20T dual-modal image sensor and an iCrest2-s edge computing device is selected as the flight platform to quickly collect dual-modal remote sensing images under various environmental conditions and perform real-time graphics processing tasks. As shown in Fig. 1(c), the forest fire detection algorithm is deployed on edge computing devices for real-time fire detection, with the dual-modal image shown in Fig. 1(b) as input. If the system detects a fire signal, the results are sent to the ground control equipment for subsequent decision-making and processing. If no fire is detected, the drone will continue to monitor the fire.

A. Lightweight Dual-Modal Fire Detection Method Architecture

The proposed detection network is based on YOLOv8 [15]. Compared with other algorithms, YOLOv8 has powerful multiscale object detection capabilities and adaptability to complex environments. In addition, its superior inference speed provides better support for real-time applications in computing-resource-constrained environments, which is critical for rapid response to forest fires. Fig. 2 shows the overall architecture of the proposed model, which is mainly composed of the backbone layer, the head layer, and the detection layer.

In the backbone layer, the single-modal input of the original YOLOv8 was changed to a dual-modal image input, and the accuracy and robustness of forest fire detection were improved by combining dual-modal remote sensing images at the same time. For dual-modal features at different scales, the DFM is designed for feature fusion, and the fusion features are fed to the head layer for further aggregation of features. Finally, the detection result is generated through the detection layer.

To better apply the proposed lightweight dual-modal forest fire detection network to UAV platforms or other edge computing devices with limited computing resources, it is necessary to

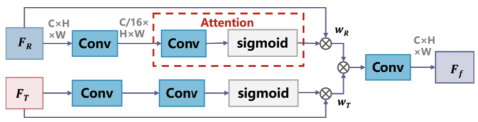

Fig. 3. Structure of the lightweight dual-modal feature fusion module.

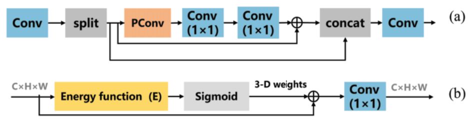

Fig. 4. Structure of (a) C2f-F and (b) SimAM.

lightweight the detection network. Therefore, the lightweight convolutional structure named C2f-F based on FasterNet [16] is introduced to replace the C2f structure in the backbone and head layer to meet the lightweight requirement. In addition, a parameter-free simple attention module (SimAM) [17] is introduced to improve model detection performance and meet lightweight designs.

B. Dual-Modal Feature Fusion

To effectively fuse dual-modal features and ensure the lightweight of the network model, this letter proposes a lightweight feature fusion module based on feature-level fusion strategy, as shown in Fig. 3.

Initially, the input features are processed using the convolution module, reducing the number of feature channels to one-sixteenth of their original size, thereby reducing the computational load of subsequent processes. Then, an SimAM is used to generate feature weights, weighting the input RGB and thermal features. The attention module consists of two parts: a convolutional layer and a sigmoid function. The weighted dual-modal features are stitched together to achieve feature fusion. Finally, another convolutional module is used to process the output fusion features to recover the number of feature channels.

C. Model Lightweighting and Attention Mechanism

Considering the limitation of hardware computing resources when using UAVs perform fire monitoring tasks, the lightweight of the detection network has become a key measure to improve its real-time performance and efficiency. To this end, the lightweight convolutional structure C2f-F is introduced to replace the original C2f, and the structures of C2f-F are shown in Fig. 4(a). C2f-F substitutes the bottleneck architecture of the original C2f with the FasterBlock structure. The FasterBlock structure is based on the FasterNet design, which aims to reduce the computational load in the feature extraction process, which comprises one partial convolution (PConv) and two Conv(1×1). PConv is based on the depthwise convolution (DWConv) design, which only applies conventional Conv on some input channels to achieve simpler and more efficient spatial feature extraction. Conv(1×1) is used to consolidate the features of individual channels while keeping the computational cost low.

TABLE I

EXPERIMENTAL CONFIGURATION

Parameters	Description
GPU	NVIDIA RTX2080-8G
CPU	Intel Core i7-9700k
RAM	16GB
Operating System	Linux Ubuntu 18.04
Development Language	Python 3.8
Deep Learning Libraries	Torch 1.9
CUDA	11.0

TABLE II

EXPERIMENTAL HYPERPARAMETER SETTINGS

Hyperparameters	Numerical value
Batch size	8
eight decay	0.0005
Epochs	100
Momentum parameter	0.937
Initial learning rate	0.01

To enhance detection accuracy while maintaining lightweight, the parameter-free attention mechanism SimAM is integrated into the backbone layer. Unlike conventional attention mechanisms, SimAM computes 3-D attention weights without introducing additional parameters, aligning with the stringent computational constraints of UAV edge devices. In addition, SimAM is grounded in spatial inhibition theory from visual neuroscience, which is able to enhance significant fire features while suppressing irrelevant background features. The structure of SimAM is shown in Fig. 4(b). The attention weights are obtained by the energy function E and the sigmoid function, the input features are weighted, and finally, the features are further processed by a Conv(1×1) to obtain the output features. Specifically, the energy function E quantifies the uniqueness of each feature by measuring its deviation from the spatial mean. Features with lower energy values are considered unique, amplified by sigmoid-based weighting, allowing the network to focus more on fire features in RGB and thermal images. The function E is expressed as follows:

$$E = \frac{(x_i - \overline{x})^2}{4\left(\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n} + \lambda\right) + 0.5} \tag{1}$$

where $\overline{x}$ denotes the mean value of x in the spatial dimension, x is the input feature, $n = W \times H - 1$ represents the effective space size of the feature, and λ represents the small constant, which is set to 10^{-4}.

III. EXPERIMENTAL RESULTS AND ANALYSIS

To evaluate the performance of the proposed method, a series of experiments are performed on the public dataset FLAME2 [18] and the fire dataset presented in [7]. Table I lists the detailed information of the experimental equipment. The experimental hyperparameters in three experiments are set as shown in Table II.

TABLE III
RESULTS OF ABLATION EXPERIMENT

Method	Datasets	AP/%	Params/MB
YOLOv8-n	R	90.6	3.0
YOLOv8-n	T	92.5	3.0
Ours(no SimAM and C2f-F)	R+T	95.3	4.3
Ours(no C2f-F)	R+T	97.8	4.3
Ours(no SimAM)	R+T	95.7	2.9
Ours	**R+T**	**98.4**	**2.9**

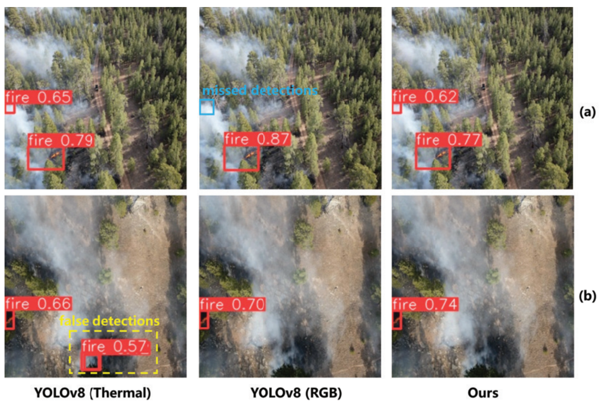

Fig. 5. Detection results on the FLAME2 dataset. (a) Using RGB image as input. (b) Using thermal image as input.

A. Dataset and Evaluation Indicators

The FLAME2 dataset comprises aerial footage recorded by a drone over a canopy pine forest in northern Arizona during a controlled burn in 2021. This dataset consists of corresponding pairs of RGB and thermal remote sensing images, with the RGB images having a resolution of 1920×1080 pixels and the thermal images measuring 640×512 pixels. A total of 5531 pairs of dual-modal frame images are extracted from the FLAME2 dataset and registered to 320×320 size. Use a tagging tool called Labelimg to mark the fire in the image.

The fire dataset proposed in [7] was acquired by a DJI M300RTK UAV with a dual-modal camera and simulated forest fires by artificially lit small fires, including RGB images and thermal images.

The average precision (AP) metric is used to assess the detection performance of the network, while the number of parameters (Params) is used to gauge the model's lightweight.

B. Ablation Experiment

To verify the effectiveness of the proposed lightweight dual-modal forest fire detection network, an ablation comparison experiment was designed. In the ablation experiment, the traditional YOLOv8 is used as the baseline model. The comparison method includes YOLOv8 with only RGB images, YOLOv8 with only thermal images, and the proposed method without the C2f-F module and the proposed method without SimAM.

The experimental results are shown in Table III, where R denotes the RGB image and T denotes the thermal image. It can be seen that the AP value of the input of the dual-modal image is better than that of the single-modality, which verifies the effectiveness of the dual-modal remote sensing image in improving the accuracy of fire identification. After adding the C2F-F module, the number of network parameters is reduced by 1.4 MB, which is 0.1 MB less than that of the single-modal YOLOv8-n. Combined with the parameter-free SimAM attention module, the AP value can be increased by 2.7%, which significantly enhances the network performance. The results show that the proposed method can achieve lightweight while ensuring the detection performance and is suitable for UAV platforms with limited computing resources.

Fig. 6. Detection results on the fire dataset proposed in [7].

Fig. 7. Edge computing device iCrest2-s.

Fig. 5 shows some of the results of the comparison experiment. When using only a thermal image as input, ash with the same temperature characteristics as fire is incorrectly identified as fire, as shown by the yellow dotted box. There are missed detections when using only RGB images, as shown in the blue box. The above results further verify the effectiveness of the proposed dual-modal forest fire detection method.

C. Generalization Experiments

To further evaluate the generalization ability of the proposed model, experiments are carried out on the fire dataset proposed in [7]. Fig. 6 shows the fire detection results of the generalization experiment, and it can be seen that the proposed model still has high detection accuracy on this dataset. The experimental results further prove that the proposed fire detection model has good generalization ability.

D. Comparative Experiment

To further validate the proposed detection method, comparative experiments were carried out with several advanced detection methods on the FLAME2 dataset. These include YOLOv5 [19], YOLOv7 [20], YOLOX-Tiny [21], YOLOv8 [15], YOLOv10-l [22], YOLOv11-l [23], YOLOv12-l [24], FTA-DETR [25], Faster R-CNN [26], CTCY [7], and Fire-DETR50 [14]. Among these methods, FTA-DETR, CTCY, and Fire-DETR50 are networks specifically designed for fire detection. In addition, the Fire-DETR50 and CTCY also use dual-modal images as inputs.

The experimental results shown in Table IV show that the lightweight dual-modal network in this letter is superior to the mainstream methods in terms of model parameters, and the AP value is second only to FTA-DETR and significantly better than the dual-modal methods Fire-DETR50 and CTCY. These results validate that our method achieves superior detection performance while maintaining light weighting.

TABLE IV
RESULTS OF COMPARATIVE EXPERIMENT

Method	Datasets	AP/%	Params/MB
YOLOv5-s	R	93.47	7.0
YOLOv7	R	93.80	36.5
YOLOX-Tiny	R	91.90	5.0
YOLOv8-n	R	90.60	3.0
YOLOv10-l	R	95.46	24.4
YOLOv11-l	R	94.33	25.3
YOLOv12-l	R	97.79	26.4
FTA-DETR	R	98.56	83.0
Faster R-CNN	R	73.29	41.3
Fire-DETR50	R+T	96.00	24.6
CTCY	R+T	97.53	17.1
Ours	**R+T**	**98.40**	**2.9**

TABLE V
iCREST2-S CONFIGURATION

Parameters	Description
GPU	NVIDIA Jeston Xavier NX
AI	supports mainstream machine learning frameworks
Memory	8GB 128-bit LPDDR4x

TABLE VI
COMPARATIVE EXPERIMENTAL RESULTS ON iCREST2-S

Method	Params/MB	AP/%	FPS
YOLOv5-s	7.0	93.47	7.0
YOLOv7	36.5	93.80	1.7
YOLOX-Tiny	5.0	91.90	4.5
YOLOv8-n	3.0	90.60	14.43
Faster R-CNN	41.3	73.29	0.7
Ours	**2.9**	**98.40**	**20.67**

E. Edge Computing Device Experiments

The proposed detection network is deployed on the edge computing device iCrest2-s for additional comparative experiments. Fig. 7 shows the iCrest2-s edge computing device used in this letter, and its hardware parameters are shown in Table V. Use AP, frames per second (FPS), and Params as performance evaluation metrics. The results shown in Table VI further demonstrate that the proposed method can accurately and real-time detect forest fires on UAV platforms.

IV. CONCLUSION

In this letter, a lightweight forest fire detection method based on dual-modal images is proposed for the UAV platform with limited computing resources. By designing a lightweight feature fusion module, the advantages of RGB image and thermal image are effectively combined to improve the accuracy of fire detection while further reducing model parameters through the introduction of a lightweight module. The experimental results indicate that the proposed method attains high detection accuracy while maintaining small model parameters in forest fire detection tasks. Moreover, on edge computing devices, the method ensures real-time and precise forest fire detection. Future research will focus on fire extinguishing strategies to improve forest fire management and control.

REFERENCES

[1] W. R. Anderegg et al., "A climate risk analysis of Earth's forests in the 21st century," *Science*, vol. 377, no. 6610, pp. 1099–1103, 2022.

[2] J. Zheng, S. Yuan, W. Li, H. Fu, L. Yu, and J. Huang, "A review of individual tree crown detection and delineation from optical remote sensing images: Current progress and future," *IEEE Geosci. Remote Sens. Mag.*, vol. 13, no. 1, pp. 209–236, Mar. 2025.

[3] L. Qiao, S. Li, Y. Zhang, and J. Yan, "Early wildfire detection and distance estimation using aerial visible-infrared images," *IEEE Trans. Ind. Electron.*, vol. 71, no. 12, pp. 16695–16705, Dec. 2024.

[4] C. Yuan, Y. M. Zhang, and Z. Liu, "A survey on technologies for automatic forest fire monitoring, detection, and fighting using unmanned aerial vehicles and remote sensing techniques," *Can. J. Forest Res.*, vol. 45, no. 7, pp. 783–792, Jul. 2015.

[5] J. Zheng et al., "Surveying coconut trees using high-resolution satellite imagery in remote atolls of the Pacific ocean," *Remote Sens. Environ.*, vol. 287, Mar. 2023, Art. no. 113485.

[6] J. Zheng et al., "Growing status observation for oil palm trees using unmanned aerial vehicle (UAV) images," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 95–121, Mar. 2021.

[7] L. Mu, Y. Yang, B. Wang, Y. Zhang, N. Feng, and X. Xie, "Edge computing-based real-time forest fire detection using UAV thermal and color images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 6760–6771, 2025.

[8] Y. Wei, Y. Liu, H. Song, and Z. Lu, "A method of rainfall detection from X-band marine radar image based on the principal component feature extracted," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[9] Y. Wang, Y. Han, Z. Tang, and P. Wang, "A fast video fire detection of irregular burning feature in fire-flame using in indoor fire sensing robots," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.

[10] Z. Qadir, K. Le, V. N. Q. Bao, and V. W. Y. Tam, "Deep learning-based intelligent post-bushfire detection using UAVs," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.

[11] L. Huang et al., "YOLO-ULNet: Ultralightweight network for real-time detection of forest fire on embedded sensing devices," *IEEE Sensors J.*, vol. 24, no. 15, pp. 25175–25185, Aug. 2024.

[12] M. Yandouzi et al., "Semantic segmentation and thermal imaging for forest fires detection and monitoring by drones," *Bull. Electr. Eng. Informat.*, vol. 13, no. 4, pp. 2784–2796, Aug. 2024.

[13] A. K. Z. R. Rahman, S. M. N. Sakif, N. Sikder, M. Masud, H. Aljuaid, and A. K. Bairagi, "Unmanned aerial vehicle assisted forest fire detection using deep convolutional neural network," *Intell. Autom. Soft Comput.*, vol. 35, no. 3, pp. 3259–3277, 2023.

[14] J. Liu, J. Song, and X. Li, "A forest fire detection method based on adaptive feature fusion module," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Aug. 2023, pp. 135–140.

[15] G. Jocher, A. Chaurasia, and J. Qiu. (2023). *Ultralytics YOLO (Version 8.0. 0)*. [Online]. Available: https://github.com/ultralytics/ultralytics

[16] J. Chen et al., "Run, don't walk: Chasing higher FLOPS for faster neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 12021–12031.

[17] L. Yang, R. Zhang, L. Li, and X. Xie, "SimAM: A simple, parameter-free attention module for convolutional neural networks," in *Proc. 38th Int. Conf. Mach. Learn.*, vol. 139, 2021, pp. 11863–11874.

[18] X. Chen et al., "Wildland fire detection and monitoring using a drone-collected RGB/IR image dataset," *IEEE Access*, vol. 10, pp. 121301–121317, 2022.

[19] G. Jocher et al., "Ultralytics/yolov5: V7.0—YOLOv5 SOTA real-time instance segmentation," Zenodo, Nov. 2022, doi: 10.5281/zenodo.7347926.

[20] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.

[21] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

[22] A. Wang et al., "YOLOv10: Real-time end-to-end object detection," in *Proc. Adv. Neural Inf. Process. Syst.*, May 2024, pp. 107984–108011.

[23] R. Khanam and M. Hussain, "YOLOv11: An overview of the key architectural enhancements," 2024, *arXiv:2410.17725*.

[24] Y. Tian, Q. Ye, and D. Doermann, "YOLOv12: Attention-centric real-time object detectors," 2025, *arXiv:2502.12524*.

[25] H. Zheng, G. Wang, D. Xiao, H. Liu, and X. Hu, "FTA-DETR: An efficient and precise fire detection framework based on an end-to-end architecture applicable to embedded platforms," *Expert Syst. Appl.*, vol. 248, Aug. 2024, Art. no. 123394.

[26] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.