ISBN: 978-981-94-1156-6
2024 the 14th International Workshop on Computer Science and Engineering (WCSE 2024)
doi: 10.18178/wcse.2024.06.007

A Flame Recognition Method Based on Data Enhancement and Yolo-v5s

Chao Huang [1], Shangyuan Li [1], Zhangcheng Yin [1], Runxia Guo [2+], Guihang Liu [3], Jiusheng Chen [2]

[1] Sino-European Institute of Aviation Engineering, Civil Aviation University of China, China

[2] School of Electronic Information and Automation, Civil Aviation University of China, China

[3] Engineering Techniques Training Center, Civil Aviation University of China, China

Abstract. To identify fires occurring in complex environments faster and locate the flames more accurately, the synthesized data enhancement methods are proposed, then a self-built flame dataset is constructed, and a flame recognition model based on Yolo-v5s is developed. The results show that after data enhancement, the false alarm rate decreases by 18.38% and the F1 score improves by 19.79%, which achieves a significant improvement in the quality of flame recognition.

Keywords: Flame Recognition; Data Enhancement; Yolo-v5s; Flame Dataset; Data Augmentation

1. Introduction

While the light of the flames helped humans away from the primitive form of life, at the same time, fire has also caused a lot of disasters worldwide. The fire poses a serious threat to the safety of human life, especially when it occurs in buildings, where it spreads rapidly and destroys life and property[1]. The fire also threatens socio-economic development and people's normal life[2]. If a rapid flame detection system was introduced to accurately identify and locate fire initiation using computer vision technology[3], and realize the alarm at the early stage of fire, more time and space could be left for the subsequent extinguishing of fires and reduce the losses caused by fires.

Compared with traditional machine vision methods, the increasing deep learning techniques have higher feature extraction efficiency. L.Jichao et al.[4] combine the single-point multi-box detector with the Mobile-Net model to design a site robot with more than 90% accuracy in flame warning for multi-frame video. Viswanatha et al.[5] propose a method that combines Yolo and Convolutional Neural Networks (CNN) to make it more effective in wide-area vision. It is easy to see that among the existing target detection algorithms, the Yolo algorithm outperforms other models because of its extremely fast detection speed.

Unlike the targets with fixed morphology, the morphology of flames changes all the time. Therefore, improving the model's ability to recognize objects with irregular morphology is crucial to enhancing the model's detection accuracy. Jin et al.[6] proposed a real-time fire smoke detection algorithm based on multi-scale feature information and an attention mechanism. Zhang et al.[7] also proposed a smoke simulation method based on a mathematical model. In addition, the size of the flame will gradually expand with the intensification of the fire, and enhancing the model's ability to recognize small-sized flames will advance the time for the fire to be detected. Z.Yuanyuan et al.[8] used a clustering algorithm to optimize a multi-scale prior frame to enhance the recognition of small-size flames in flame detection based on Yolo-v3. Bari[9] and J.Wenping et al.[10] also went through multiple migration learning to improve the detection of initial fires.

These studies focus on optimizing the model based on the characteristics of the flames and give less consideration to real-life situations under the surveillance field of view. In many cases, the monitor employs a top-down view, whereas existing flame datasets predominantly feature a horizontal view or are generated by extracting frames from fire-related videos, exhibiting limited generalization performance. Therefore,

+ Corresponding author. Tel.: + 86 13612182112
E-mail address: rxguoblp@163.com.

adjustments to the dataset are necessary. At the same time, the flames are very easy to confuse with sunlight and lights at lower resolutions, and false alarm rates are higher in bright environments.

To address the above issues, we propose a flame recognition method based on data enhancement and Yolo-v5s. First, we establish a flame dataset by synthesized various data enhancement methods, then develop a fast and effective flame recognition method under the framework of the Yolo-v5s algorithm. Finally, the feasibility of the proposed method is verified through the self-constructed dataset, realizing accurate flame recognition under the monitoring perspective.

2. The Yolo-v5s Structure

Yolo-v5s is a deep learning-based target detection algorithm proposed by Ultralytics LLC in 2020[11], which can quickly and efficiently detect multiple targets in images or videos[12], and label their locations and categories. Its image detection speed is about 140 frames per second, which is much faster than traditional object detection algorithms. The model file size of Yolo-v5s is only 14 MB and the network structure of the Yolo-v5s model[13] is shown schematically in Fig. 1.

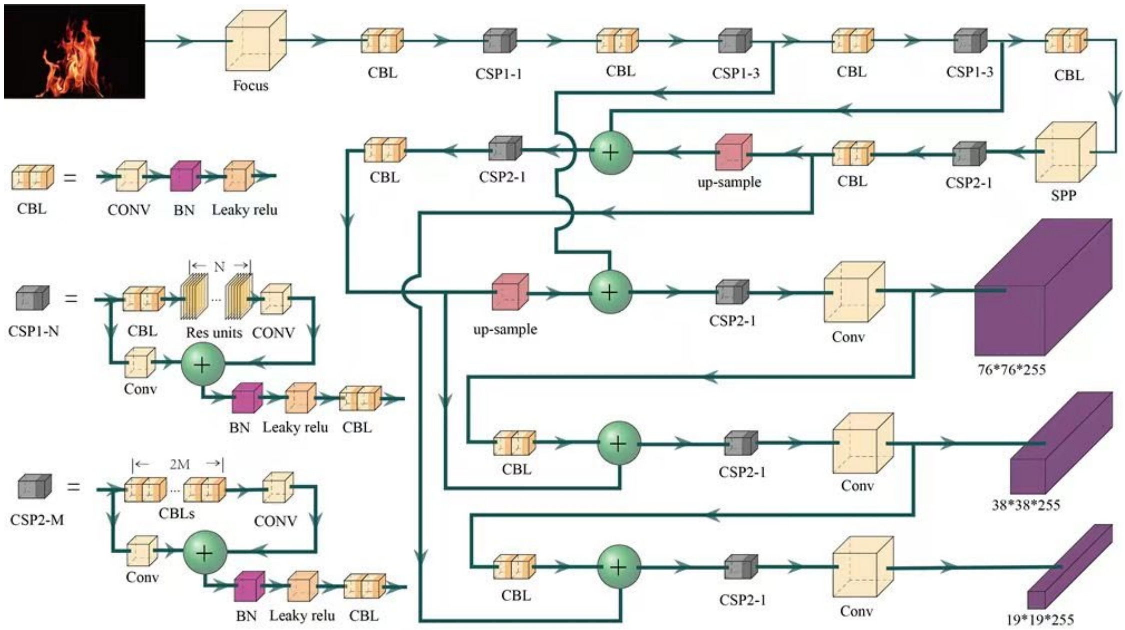

Fig. 1: Network structure diagram of Yolo-v5s

3. Data Enhancement Methodology

The training sets of common flame recognition models are all flame images, which make the model have a low miss rate in flame recognition, but the distribution of flames in the dataset is not uniform and the generalization ability is poor. Therefore, in the study, the generalization performance of the dataset will be optimized through a series of data enhancement methods to reduce the false positives of the model and increase the accuracy.

3.1. Negative Sample Screening Method

During the training of the prediction model, it was found that negative samples have a great impact on the accuracy of the trained model. If pictures with bright, reddish-yellow elements are fed into the model without data augmentation, there is a high probability that the model will misreport them as flame pictures. As shown in Fig. 2, the evening sun, sunrise, incandescent lights, neon lights, and other non-flame pictures will be misidentified as flames by the model, which can be cleverly utilized to provide negative samples used for subsequent training as a way of reducing the false alarm rate of the model.

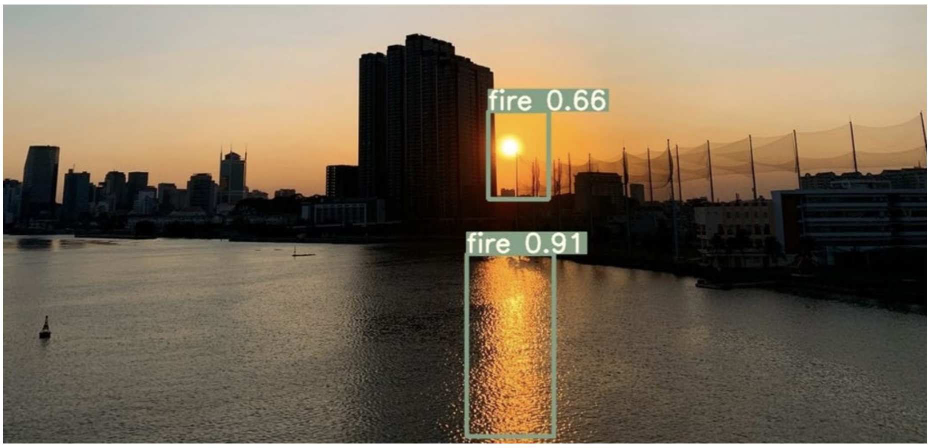

Fig. 2: Non-flame image incorrectly identified as a flame

In flame recognition, confidence refers to the probability that the model presumes an area to be a flame, usually expressed as a numerical value ranging from 0 to 1. The greater the confidence level, the greater the likelihood that the model believes the area is a flame. To ensure the accuracy of the detection results, it is necessary to set a confidence threshold, when the confidence level is greater than this threshold, the model will determine that the region appears to be a flame, otherwise, it is considered that the region contains no flame.

In the flame recognition model based on the Yolo-v5s algorithm, the confidence level of the flame ($\mathrm{P_r(Fire) \cdot IoU}$) equals to the product of the probability of the flame category appearing in the detection frame($\mathrm{P_r(Fire|Object)}$) and the frame confidence score($\mathrm{P_r(Object) \cdot IoU}$), i.e. :

$$P_r(Fire) * IoU_{pred}^{truth} = P_r(Fire \mid Object) * P_r(Object) * IoU_{pred}^{truth} \tag{1}$$

Where $P_r(Fire)$ denotes the probability that the object presents in the detection frame is of flame type, $P_r(Fire|Object)$ denotes the probability that the object is of flame type under the condition that the object is present in the detection frame, $P_r(Object)$ represents the probability that the object is present in the detection frame, and IoU is the abbreviation of Intersection over Union, i.e., the degree of overlap between the prediction frame (predicted box) and the real frame (ground truth):

$$IoU = \frac{A \cap B}{A \cup B} \tag{2}$$

Where A denotes the prediction box and B denotes the ground truth.

The collected flame pictures as well as non-flame pictures with similar features to the flame were input into the model in batches. After repeated experiments to verify the results, for the output results when inputting flame pictures, the data with a confidence level above 0.6 were screened to be added into the flame dataset; for the output results when inputting non-flame pictures, all the mislabels misidentified as flames were changed to non-flame labels in batches and were added into the next stage of the dataset as negative sample data.

3.2. Boundary Extension Cropping Method

To expand the dataset of small samples, the most common method used in existing studies is to randomly crop the images, which may cause valid information to be removed or partially lost. Therefore, we propose a boundary-extended cropping method to turn the original multi-target image into a single-target image containing sufficient background information.

The specific implementation process is illustrated in Figure 3: First, reading label information from images, iteratively processing each labeled bounding box, then generating cropped images equal to the original image's label count. As shown in Fig.4, when a label is calculated, other label boxes need to be considered as obstacles, starting from the label box boundaries, expanding the boundaries in four directions, up, down, left and right, until it touches the obstacle or image boundary.

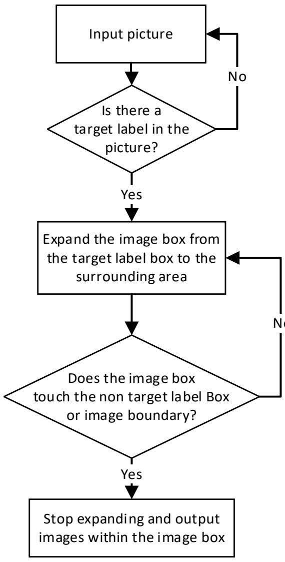

Fig. 3: Edge extension clipping method

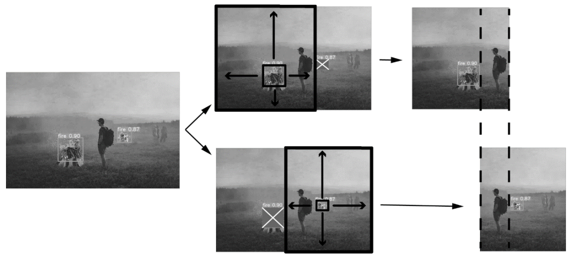

Fig. 4: Edge extension clipping method diagram

The advantages of this method are mainly reflected in the expansion of the small sample data set while avoiding the loss of effective information that may be caused by random cropping. In addition, when using this method to evaluate the accuracy of a specific label, other non-target label boxes will be regarded as obstacles, thus ensuring the independence between different targets and helping to improve the generalization performance of the model.

3.3. Geometric Transformation

Considering the different morphology of the flame captured by the lens from different directions, in order to simulate different viewpoints and scenes, and to improve the recognition ability of objects captured from opposite and side directions[14]. The position of each pixel point and marker point of the image after rotation is calculated as follows:

Assume that the image origin is located at the upper left and set the image coordinates of the rotation point M to (x, y) and the image center coordinates to (x_c, y_c). After moving the origin to the center of the image, the coordinate of point M becomes $(x - x_c, y - y_c)$. If the point M is rotated by an angle of a, the coordinates of the rotated point (x', y') in the coordinate system of the center of the image are:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(a) & -\sin(a) \\ \sin(a) & \cos(a) \end{bmatrix} * \begin{bmatrix} x - x_c \\ y_c - y \end{bmatrix} \tag{3}$$

Considering the characteristics of the flame, when flipping, it is done only in the horizontal direction. Due to the variable shape of the flame, the image can be segmentally affixed; since the viewpoint of the surveillance is often top-down, the image can be perspective transformed. The effect of some of the geometric transformation methods is shown in Fig. 5.

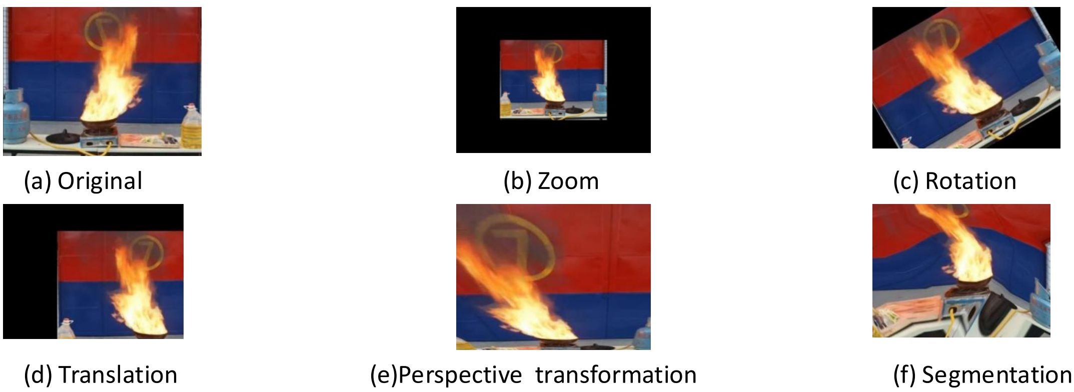

Fig. 5: Schematic diagram of geometric transformation methods

3.4. Smoke Simulation Method

Some fires occurring indoors often produce smoke that is difficult to diffuse, which often blurs the view under surveillance quickly, resulting in pixel-level feature-based information that is not obvious, making recognition difficult. In this study, a real-time fire smoke detection algorithm is used to simulate the visibility degradation of smoke from fire by applying random smoke, Gaussian blurring, and contrast enhancement to the image in order to improve the model's ability to recognize flames in thick smoke.

After several tests, the best model parameters in the experiment are set as in Table 1, and the effect is shown in Fig. 6.

Table 1: Training Set and Test Set Model Parameter Settings

Parameter settings	The average intensity of clouds	Frequency noise index	Standard deviation of Gaussian distribution	Minimum mixed alpha
Training set	251.8	-1.4	2	0.7
Test set	254.2	0.8	2	0.8

Parameter settings	Frequency index	Alpha sampling multiple	The sparsity index of clouds	Density multiplier
Training set	-2	0.3	0.9	0.5
Test set	-1	0.3	0.9	0.6

(a) Adjust contrast (b) Gaussian blur (c) Simulate smoke alone

Fig. 6: Schematic diagram of smoke simulation method

As shown in Fig.7, a random combination of the geometric transformation method and the smoke simulation method using a similar data augmentation method lead to a multiplicative augmentation of the original samples based on the characteristics of the fire flames and a better simulation of the diversity of the fire occurrence scenarios.

(a)Geometric transformations (b) Integrated simulation of smoke (c) Combined effects

Fig. 7: Combination of geometric transformation and smoke simulation method

4. Experiments

4.1. Experimental Platform

The experimental system environment is: Intel i7-10750H processor, 16G RAM, NVIDIA GeForce RTX 2060, using a development environment that includes python version 3.9.8, pytorch version 1.11.0, cuda version 11.1, and cudnn version 8.1.

The training parameters were adjusted by setting the number of images in each batch to 2, the number of training rounds to 500, the initial learning rate to 0.01, the number of warm-up rounds to 3, the IOU threshold for training to 0.2, and the IOU threshold for validation to 0.5.

4.2. Recognition Model Training

The training of the model requires labelled images, and the model will automatically extract image features from the images in the dataset according to the labelled regions. The labelled file is a txt document, each line of information in the document is label information, and each label information contains the label category, the label midpoint position, and the width and height of the label box. In order to obtain the labelled images, 1000 randomly selected images from the fire images were manually labelled. These images cover the behavior of flames under different environmental conditions, including day and night, indoor and outdoor settings, as well as various lighting conditions. Labelimg is a tool commonly used in the field of deep learning for data annotation, which was used to complete the labelling of these images.

By visiting the image resource website and the database collected from the public domain of BILKENT University[15], 8972 flame images were screened, out of which 8000 flame images were taken as the training set and the remaining 972 images were taken as the test set and the control model was trained.

4.3. Data Enhancement

After testing, it was found that although it has a good performance in simple flame recognition, the anti-interference ability is extremely poor, if the input to the system is bright, reddish-yellow objects will be the probability of the model will be incorrectly recognized as a flame. By using the negative sample screening method, 4987 non-flame images with similar characteristics to the flame were screened from the gallery, and 4000 of them were added to the training set, and the remaining 987 were added to the test set, so as to improve the model's generalization ability and accuracy. In the data enhancement stage, a total of 10 enhancement methods are set up, including boundary extended cropping method, segmented affine, zoom-in and zoom-out, horizontal flip, rotation, random four-point perspective transformation, displacement, Gaussian blur, smoke, and adjusting contrast, and the intensity of any enhancement method will be randomized within a certain interval.

When augmenting the pictures in the test set, new pictures with more than five augmentation methods were randomly generated. To ensure that the number of pictures in the training set of the experimental group is the same as that of in the comparison group, another 4,000 flame pictures are randomly selected from the existing training set and mixed with 4,000 non-flame pictures to form the training set of the experimental group, totally 8,000 pictures.

4.4. Analysis of Experimental Results

The recognition quality of the model is evaluated by Precision, Recall, and the F1-score, which are defined below:

$$\text{Precision}=\frac{\text{TP}}{\text{TP+FP}} \tag{4}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP+FN}} \tag{5}$$

$$\text{F1 score} = 2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \tag{6}$$

Where TP (True positive) is the number of flame pictures that were correctly identified, FN (False negative) is the number of non-flame pictures that misreported flames, TN (True negative) is the number of non-flame pictures that were correctly processed, and FP (False positive) is the number of flame pictures that should have been reported but were not.

Based on the comparison model, the model trained using the enhanced data set is the experimental model. In order to evaluate the effect of data enhancement, the test set was tested to assess the model quality using the comparison model and the experimental model at a confidence threshold in steps of 0.1, and from 0.2 to 0.8, respectively. The TP and FP curves for the comparison and experimental groups are shown in Fig. 8 and Fig. 9, respectively. As it can be seen from the figures, our proposed model has a significant improvement over comparison model at all confidence thresholds.

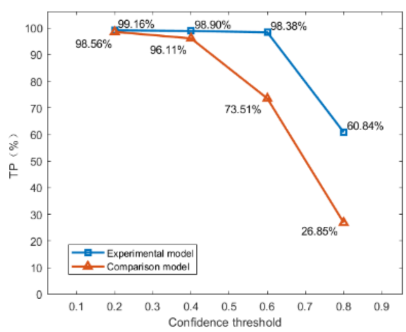

Fig. 8: Confidence threshold and TP values

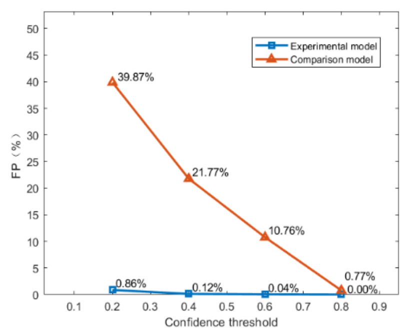

Fig.9: Confidence threshold and FP values

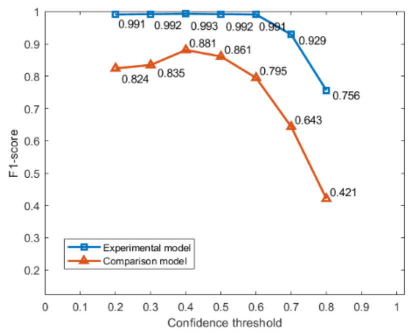

Fig.10: Confidence threshold and F1 score

The F1 score is the reconciled mean of precision and recall, which is a better measure than precision or recall alone. The change curves of the F1 score for the control and experimental groups are shown in Figure 10, and the overall validation results are shown in Table 2. Our proposed model also achieves better performance.

Table 2: Results of Comparison Model and Experimental Model at Different Confidence Threshold

Confidence threshold		0.2	0.3	0.4	0.5	0.6	0.7	0.8
Comparison model	TP (%)	98.56	98.15	96.09	88.58	73.15	49.38	26.85
	FP (%)	39.92	36.47	21.68	16.92	10.74	4.05	0.81
	Precision (%)	70.86	72.60	81.36	83.75	87.03	92.31	97.03
	Recall (%)	98.56	98.15	96.09	88.58	73.15	49.38	26.85
	F1-score (%)	82.44	83.46	88.11	86.10	79.49	64.34	42.06
Experimental model	TP (%)	99.07	99.07	98.87	98.56	98.35	86.83	60.80
	FP (%)	0.81	0.61	0.20	0.10	0.10	0.10	0.00
	Precision (%)	99.18	99.38	99.79	99.90	99.90	99.88	100.00
	Recall (%)	99.07	99.07	98.87	98.56	98.35	86.83	60.80
	F1-score (%)	99.13	99.23	99.33	99.22	99.12	92.90	75.62

After data enhancement, the confidence level of the model recognizing the flame images is significantly improved, and the false alarm rate of the performance model for non-flame images is significantly decreased, with an average decrease of 18.38%; when the confidence threshold is set to 0.7, the F1 score of the control group has been decreased to 79.4%, while the experimental group still has 92.9%, with an average improvement of 19.79% in F1-score. We draw the conclusion that under the premise of keeping the model architecture unchanged, the application of data augmentation methods effectively expands the number of flame samples, thereby enhancing the accuracy of flame detection.

The detection speed of several mainstream target detection algorithms on various datasets, and the performance metric are shown in Table 3. The results show that the Yolo-v5s achieves the best performance in F1 score. Besides, the model also exhibits strong generality in common environments such as indoor settings, forested areas, and factories.

Table 3: Comparison of detection rate and accuracy of several target detection algorithms

Algorithm	Dataset name	Detection speed	F1 score
Yolo-v5s	Self-built dataset	166 fps	92.9%
Faster R-CNN	VOC2012 dataset	45 fps	91.3%
SSD	COCO dataset	58 fps	89.2%

5. Conclusion

After data enhancement by the proposed methods, a new flame dataset is built, then the false alarm rate is significantly reduced, and the quality of flame recognition by the flame detection model is greatly improved. The research focuses on the current use of frame-by-frame detection of video, and how to utilize the temporal information of video images, which need to be studied in depth in the future.

6. Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (62173331), the Fundamental Research Funds for the Central Universities (3122019181).

7. References

[1] S. K. Akagi, R. J. Yokelson, C. Wiedinmyer, et al. Emission factors for open and domestic biomass burning for use in atmospheric models, Atmospheric Chem. Phys. 2011, 11(9): 4039–4072.

[2] W. Permar, Q. Wang, V. Selimovic, et al. Emissions of Trace Organic Gases From Western U.S. Wildfires Based on WE-CAN Aircraft Measurements, J. Geophys. Res. Atmospheres. 2021, 126(11): e2020JD033838.

[3] Y. Li, Q. Zhang, Z. Shen, B. Zuo. Flame saliency detection based on FMF, China Saf. Sci. J. 2019, 29(5): 56-61.

[4] J. Li, S. Guo, L. Kong, Y. Yuan, Y. Sun, Design and application of a fire detection and early warning robot in construction site, China Saf. Sci. J. 2021, 31(04): 141-146.

[5] V. V, C. R. K, R. A. C. Real Time Object Detection System with YOLO and CNN Models: A Review. arXiv, 2022.

[6] C. Jin, A. Zheng, Z. Wu, C. Tong. Real-Time Fire Smoke Detection Method Combining a Self-Attention Mechanism and Radial Multi-Scale Feature Connection, Sensors. 2023, 23(6): 3358.

[7] Z. Zhang, C. Yang, Y. Zhang, H. Zhu. Dynamic modeling method for infrared smoke based on enhanced discrete phase model, Infrared Phys. Technol. 2018, 89: 315–324.

[8] Y. Zhao, J. Zhu, Y. Xie, et al. A Real-Time Video Flame Detection Algorithm Based on Improved Yolo-v3, Geomat. Inf. Sci. Wuhan Univ. 2021, 46(03): 326-334.

[9] A. Bari, T. Saini, A. Kumar. Fire Detection Using Deep Transfer Learning on Surveillance Videos, Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021.

[10] W. Jiang, Z. Jiang. Research on early fire detection of Yolo V5 based on multiple transfer learning, Fire Sci. Technol. 2021, 40(01): 109-112.

[11] R. J. Yang, W. F. Li, X. N. Shang, et al. KPE-YOLOv5: An Improved Small Target Detection Algorithm Based on YOLOv5, Electronics. 2022,12(4): 817.

[12] A. G. Howard, M. Zhu, B. Chen, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv, 2017.

[13] G. Chen, H. Zhou, Z.Y. Li. Multi-Scale Forest Fire Recognition Model Based on Improved YOLOv5s, Forests. 2023,14(2): 315.

[14] Z. A. Khan, T. Hussain, FUM. Ullah, et al. Randomly Initialized CNN with Densely Connected Stacked Autoencoder for Efficient Fire Detection, Eng. Appl. Artif. Intel. 2022, 116: 105403.

[15] A. E. Gunawaardena, R. M. M. Ruwanthika, A. Jayasekara, Computer vision based fire alarming system, Moratuwa Engineering Research Conference (MERCon), IEEE, 2016.