Fire Detection Algorithm Based on an Improved Strategy of YOLOv5 and Flame Threshold Segmentation

Yuchen Zhao, Shulei Wu*, Yaoru Wang, Huandong Chen*, Xianyao Zhang and Hongwei Zhao

School of Information Science and Technology, Hainan Normal University, Haikou, 571158, China
*Corresponding Authors: Shulei Wu. Email: 595615374@qq.com; Huandong Chen. Email: chd@hainnu.edu.cn

Abstract: Due to the rapid growth and spread of fire, it poses a major threat to human life and property. Timely use of fire detection technology can reduce disaster losses. The traditional threshold segmentation method is unstable, and the flame recognition methods of deep learning require a large amount of labeled data for training. In order to solve these problems, this paper proposes a new method combining You Only Look Once version 5 (YOLOv5) network model and improved flame segmentation algorithm. On the basis of the traditional color space threshold segmentation method, the original segmentation threshold is replaced by the proportion threshold, and the characteristic information of the flame is maximally retained. In the YOLOv5 network model, the training module is set by combining the ideas of Bootstrapping and cross validation, and the data distribution of YOLOv5 network training is adjusted. At the same time, the feature information after segmentation is added to the data set. Different from the training method that uses large-scale data sets for model training, the proposed method trains the model on the basis of a small data set, and achieves better model detection results, and the detection accuracy of the model in the validation set reaches 0.96. Experimental results show that the proposed method can detect flame features with faster speed and higher accuracy compared with the original method.

Keywords: YOLOv5; fire safety; deep learning; flame detection

1 Introduction

With the development and progress of human society, fire has become a frequent problem that cannot be ignored. Fire will pose a great threat to the safety of people's lives and property and the stability of society [1]. Therefore, timely detection in the early stage of fire can reduce the losses. In order to reduce the number of fire-related injuries, a multitude of early fire detection methods has been proposed [2,3]. Most of the early fire detection technologies are based on sensors, which are easy to be affected by external factors and have great limitations [4–7]. With the development of computer vision technology, detection technology based on video and image has brought great hope for fire detection [8–10].

In recent years, many experts and scholars have gradually deepened their research on flame recognition and achieved good results [11–14]. In flame image segmentation, the most commonly used methods are segmentation based on color space and image difference [15]. In the segmentation methods based on color space, the threshold segmentation method based on the gray histogram is the most commonly used method. Lenka et al. [16] proposed a segmentation method based on an unsupervised histogram, which calculated the threshold required for segmentation by evaluating the peak value of the image histogram. Haifeng et al. [17] proposed an anti-noise Otsu image segmentation method based on the second-generation wavelet transform, which calculated the maximum interclass variance after the noise of the target and background was suppressed in the wavelet domain through the second-generation wavelet transform, to improve the efficiency and noise resistance of the traditional algorithm. However, the threshold segmentation method requires a large amount of calculation, which cannot be adapted to a variety of occasions, and it is challenging to meet the real-time requirements of flame detection. Based on conventional color space recognition, Seo et al. [18] proposed a color segmentation method combining background subtraction, whose recognition effect has been further improved in the essential color space but is poor in flame images with complex backgrounds. Lights, sunlight, reflected light, and red objects in complex scenes will affect the accuracy of flame recognition. After foreground extraction, it often cooperates with a clustering or morphological algorithm to optimize [19]. Although it can improve accuracy, it is time-consuming to process the image, necessitates high-end hardware, and the recognition accuracy is unstable.

In addition, many scholars try to perform flame segmentation on the basis of different color spaces and integrate the segmentation effects of other color spaces. Wilson et al. [20] suggested a Video Surveillance Fire Detection System with a high detection rate and a low number of false alarms. Color models such as Hue-Saturation-Value (HSV) and YCbCr are used in the system to improve the flame segmentation effect, and the experimental results are better than the separate Red-Green-Blue (RGB), HSV, and YCbCr spatial segmentation results. The method proposed by Samuel et al. [21] combined the unique color and texture features of the flame, improved on the basis of HSV and YCbCr color spaces, and segmented the flame region gray level co-occurrence matrix (GLCM) from the image to calculate the five texture features. The method was tested on a set of fire and non-fire images, achieving the effect of a higher detection rate and a lower false positive rate. Thepade et al. [22] proposed flame detection based on Kekre's LUV color space and used the flickering characteristics of fire to confirm the existence of fire. The flame segmentation effect that combines multiple color spaces has significant advantages compared to the segmentation effect of a single color space. However, there are still some problems: it takes a long time to process images and has low processing accuracy. It will produce poor results, especially with images of small fires, thick smoke, or more background elements.

The image difference method is characterized by simple algorithm and high operation efficiency, so it is suitable for most occasions where the real-time performance of flame monitoring is relatively high. Typical differential image methods include the frame-difference method [23] and the background subtraction method. Cai et al. used the improved five-frame difference method for analysis on the basis of single detection and dispatch, [24] making the analysis results more comprehensive. Shi et al. [25] proposed a method of processing the false fire area recovered by image difference and Gaussian mixture modeling and the fire and flame candidate areas determined by color filters. Although the effect of the image difference method is generally better than the recognition effect based on the color space, the fire background images are actually diverse and difficult to be universal, and the recognition effect will be unstable in different scenes. and it is more dependent on hardware settings. The processing speed requirements are also more stringent.

Image recognition technology has been developed and applied in many fields, such as target detection, face recognition, medical image recognition, remote sensing image classification [26–29]. Literature [30,31] used a neural network model for target detection, which has high recognition accuracy and fast recognition speed. The problem of fire recognition and early warning using two-dimensional image neural network target detection classification and recognition method. This approach can make good use of the advantages of computer vision to make up for the shortcomings of traditional flame warning methods [32]. Muhammad et al. [33] used the convolutional neural network model of GoogleNet architecture for flame detection, and the speed and accuracy of flame detection obtained by experiments were improved compared with traditional methods. Chaoxia et al. [34] made two improvements to the original Faster R-CNN: the color-guided anchoring strategy and the global information-guided flame detection. Chen et al. [35] designed an image-based fire alarm system using a convolutional neural network (CNN) method for fire identification and created a system with 92% accuracy. However, using neural networks for image detection still has problems, such as a high false positive rate, a long training time, and a large scale of training-dependent datasets.

In order to solve the above problems, bootstrapping and cross-validation [36,37] are introduced in the YOLOv5 network model and the traditional color space threshold segmentation method is improved, which is replaced by the proportional threshold value [38]. The YOLOv5 network model is combined with the improved flame segmentation algorithm to classify and train flame images [39]. The main contribution of this study has threefold: first, bootstrapping and cross-validation are introduced into the YOLOv5 network model; second, it is improved on the basis of the traditional threshold segmentation method, which is replaced by proportional threshold; finally, the YOLOv5 network model is combined with the improved flame segmentation algorithm to classify and train flame images.

The rest of this paper is organized as follows. Section 2 introduces the improvement of flame threshold segmentation method. Section 3 details Improvement Based on YOLOv5 Network Model. Section 4 shows the extensive experiments and analysis on our proposal. We conclude in Section 5.

2 Improvement of Flame Threshold Segmentation Method

This section explains the implementation strategy of the proposed model. The flowchart of the method is shown in Fig. 1.

2.1 Classical Flame Threshold Segmentation Method

This section focuses on the color space flame recognition problem. In the RGB color space, according to the principle of three primary colors, the amount of light is represented by the unit of direct color light, and the expression of different colors is formed by the addition and mixing of different components of R, G, and B. Flame images in RGB space often show a large proportion of red components. Therefore, there is the following basic criteria:

$$\begin{cases} R > R_{ave} \\ G > G_{ave} \\ R > G > B \end{cases} \qquad (1)$$

Among them, R_{ave} is the mean of the red channel of all pixels in the image, and G_{ave} is the mean of the green channel. The traditional threshold segmentation method only considers the color information of the flame itself in the image. Usually, the flame image has the characteristic that the

value of the R channel is greater than the values of the G and B channels, but not all the desired images satisfy the B < G rule. For example, the color characteristics of open fire and dark fire sometimes violate this criterion. When the flame is burning, analyze the RGB color channel values in the flame image. The histogram is shown in Figs. 2a and 2b:
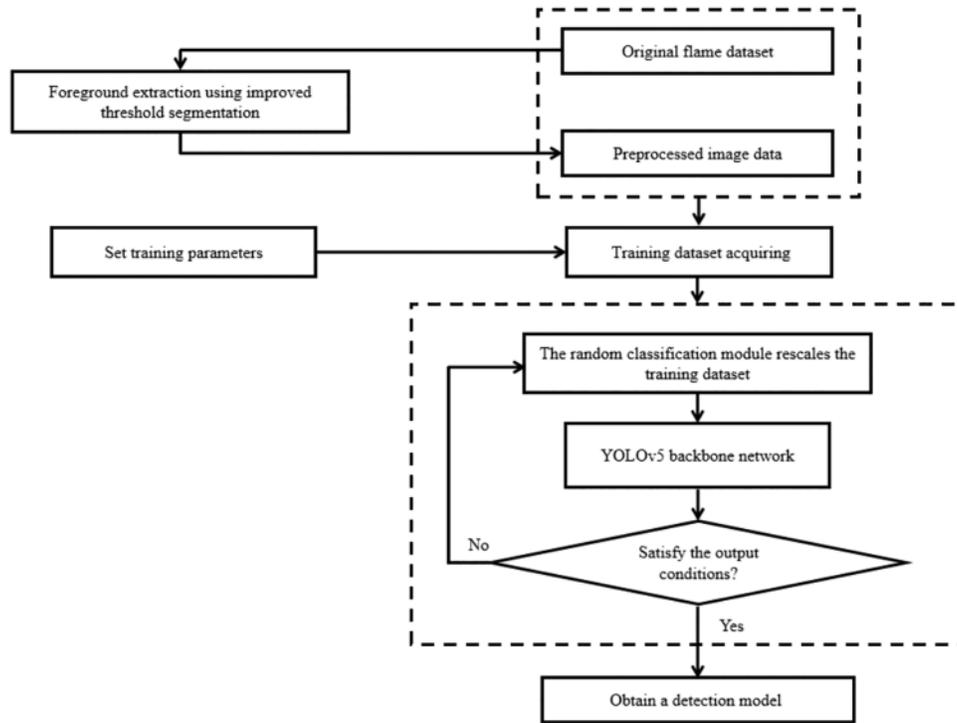

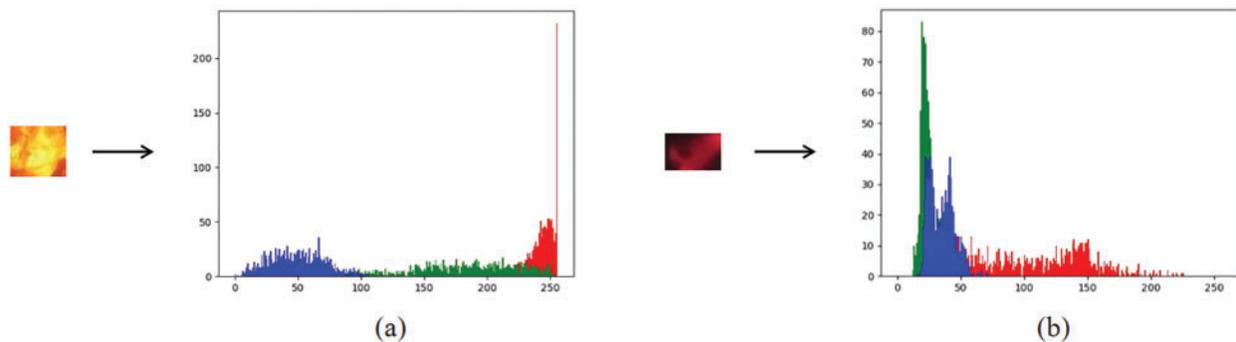**Figure 1:** Flowchart of the proposed method of this study

Figure 2: Histogram of RGB channels of open flame and dark flame (a) open flame (b) dark flame

It can be seen that, regardless of whether it is a dark fire or an open fire, the color value of the RGB channel shows a high proportion of red, but not all feature points satisfy the threshold of the G channel, which is greater than the threshold of the B channel. In the images of the open fire with more intense burning, the color characteristics of G > B are often shown, and in the dark fire parts with less severe burning, the color characteristics of G > B are sometimes shown.

Taking the background as grass, smoke, sky, and debris, the flame background partial image is used for histogram analysis, as shown in Fig. 3:
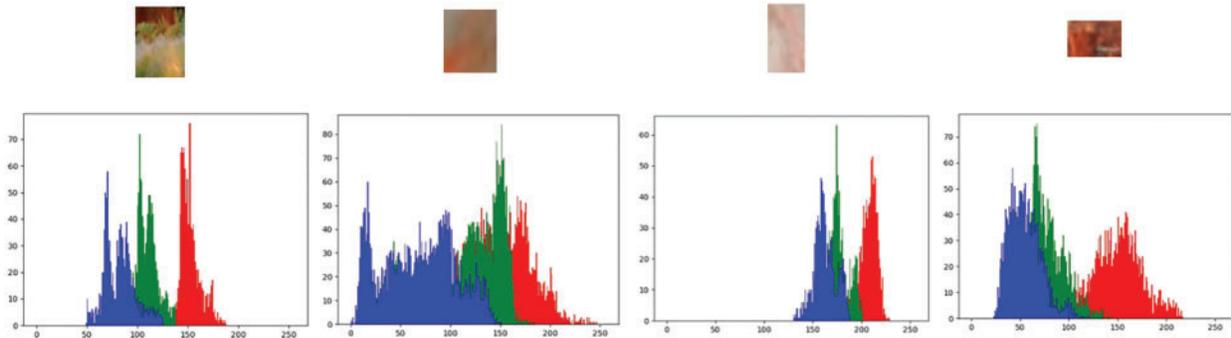

Figure 3: Comparison of flame RGB histograms under different backgrounds

In addition to the above two cases, the traditional flame segmentation model will also fail to deal with severe and full combustion images. In the flame segmentation results based on color space, the flame image of the flame core has the highest misjudgment rate. Since the flame core is the part where the flame burns most fully and violently, its color characteristics are often not red or yellow but close to white, and the RGB channel values of the flame core no longer have obvious distribution rules. Fig. 4 highlights the issue:
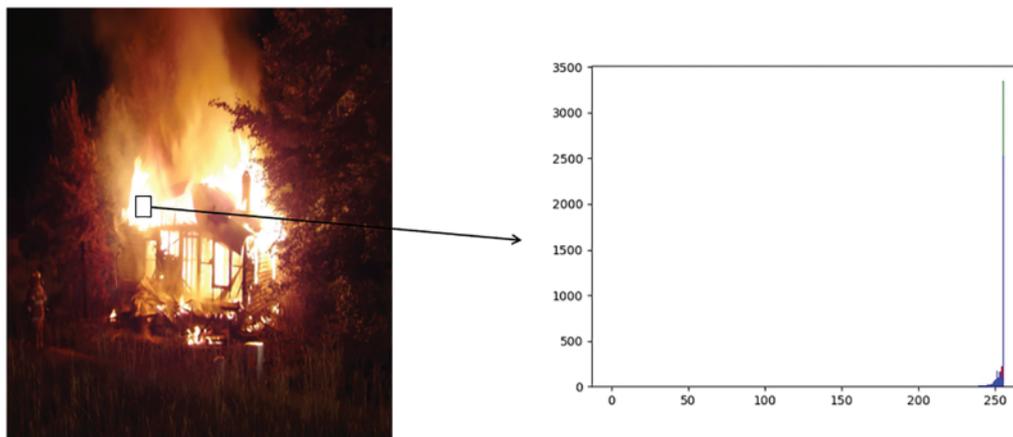

Figure 4: RGB histogram of the flame in the highlighted flame core

To sum up, the settings of the average of R_{ave}, G_{ave}, and B_{ave} in the RGB flame division rules do not have a universal application value.

2.2 Improvement Strategy

In Section 2.1, analysis of the color space flame recognition problem, although different flame background images have various performances, the values of each flame channel often show different performances in a different light and dark scenes. Cai et al. [24] proposed a classic flame segmentation

algorithm based on a large number of flame images from Celike statistics, improved the algorithm of the RGB color space, offered a more accurate flame segmentation rule applicable to the video flame recognition method, and changed the difference between different color channels. The segmentation method introduces the scheme of varying color channel ratios for analysis and has achieved good results. This paper makes further improvements on the improvement ideas of the literature, trying to expand the threshold analysis of a single channel to the threshold value of the three combination ratios of RG, GB, and RB to analyze. The outer flame is analyzed separately, and a foreground extraction rule that can preserve the flame characteristics to the greatest extent is proposed as formula (2):

$$
\begin{cases}
\dfrac{R(x,y)}{B(x,y)} \geq T_{RB} \\[2mm]
\dfrac{B(x,y)}{G(x,y)} \geq T_{BG} \\[2mm]
\dfrac{R(x,y)}{G(x,y)} \geq T_{RG}
\end{cases}
\tag{2}
$$

As mentioned above, the color characteristics of the flame will fail in the very intense flame core part, and its RGB channel value is above 230. The outer flame part, with a lower temperature during combustion, has lower RGB channel values to preserve the flame characteristics as much as possible. Based on the actual threshold, it is necessary to add the restoration of the high-brightness area and the low-dark site in the color space to ensure that the flame features are not removed in the flame recognition. For the flame image in the highlighted area, due to the high temperature, the color of the RGB channel is close to white, the channel values are all high, and the difference is small. After a large number of experimental comparisons, formula (3) is added on the basis of formula (2) to restore the highlighted area.

$$
R(x,y) \geq T_R
\tag{3}
$$

The flame image processing for low-dark parts differs from that of highlight parts. In a darker flame, although the RGB channel value is smaller, the red and yellow channel values are more weighted, and its color fluctuation is mainly disturbed by the background information. Therefore, when restoring the flame area in the low-dark area, the color value of each channel should be comprehensively considered to determine the restoration threshold. After experimental comparison, the formula is defined as follows formula (4):

$$
\begin{cases}
Gray = R_w * R(x,y) + G_w * G(x,y) + B_w * B(x,y) \\[2mm]
Gray \geq T_G
\end{cases}
\tag{4}
$$

3 Improvement Based on YOLOv5 Network Model

3.1 YOLOv5 Network Structure

YOLO [40] is one of the mainstream algorithms in the field of deep learning today, and it is a one-stage, single-step detection method based on the regression method of deep learning. Although the detection accuracy is not as good as that of Mask-RCNN and other network systems, it also has higher detection accuracy. Its simple and fast recognition efficiency makes it more suitable for real-time detection. In fire monitoring work based on computer vision, detection algorithms are needed

to provide real-time detection capabilities. Therefore, it is of great practical significance to study how to combine YOLOv5 with flame recognition to improve detection accuracy. Fig. 5 is the YOLOv5 architecture.
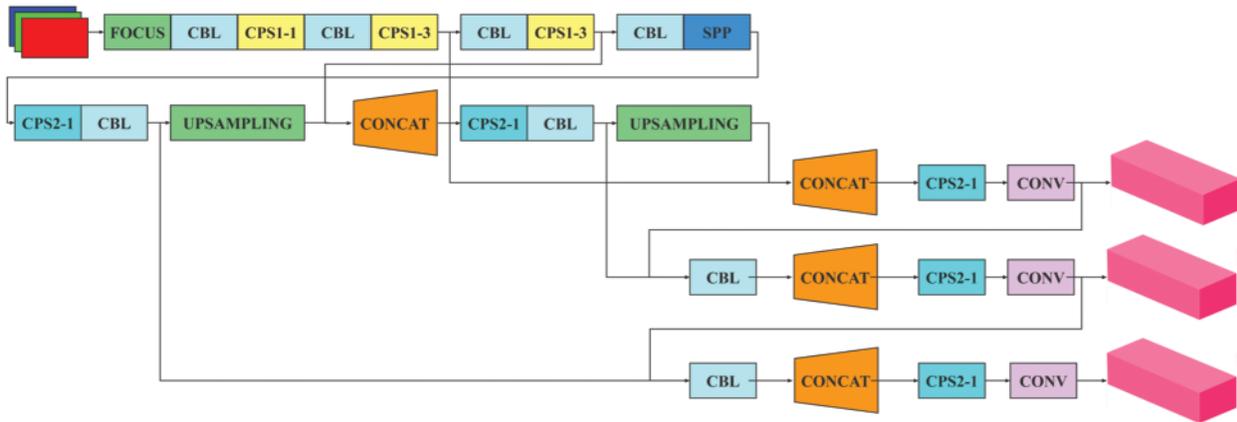

Figure 5: Schematic diagram of YOLOv5 [40]

3.2 Randomized Cross-Validation Based on Bootstrap Method

In the recognition work of some neural networks, it is often impossible to obtain large network datasets for training. In the position of classification training on small sample data sets, the training effect is often prone to the unstable impacts. When the population from which the sample comes cannot be described by a normal distribution, using the bootstrap method to analyze the data can make the data training effect better [41,42]. The core idea of the bootstrap method is to use random replaceable sampling detection, which has the effect of adjusting the data in the training process.

In the target detection work using YOLOv5, the network validation set plays the function of cross-validation. Based on using the original YOLOv5 neural network for target detection and classification training, this paper introduces the idea of the bootstrap to process the input data of the neural network and the verification data, including random extraction of smoke segmentation images, random extraction of flame segmentation images, and random extraction of original dataset images [43,44]. Before the random extraction process of the original image by the bootstrap, the original image database is preliminarily divided, among which 269 images are images mainly containing flame data, which are also used as training data sets, and 206 are informational images containing positive and negative samples. The specific data set division method is shown in Fig. 6:

After dividing the data set, we add the random classification module to the training network process. After each round of training, the network will automatically update the training data set, and after training the updated training set, new weight information will be obtained, and the network will also adjust according to the new calculation weight. We process the training dataset in this way resulting in random cross-validation of the network. The flow chart for the above verification idea is shown in Fig. 7 below.
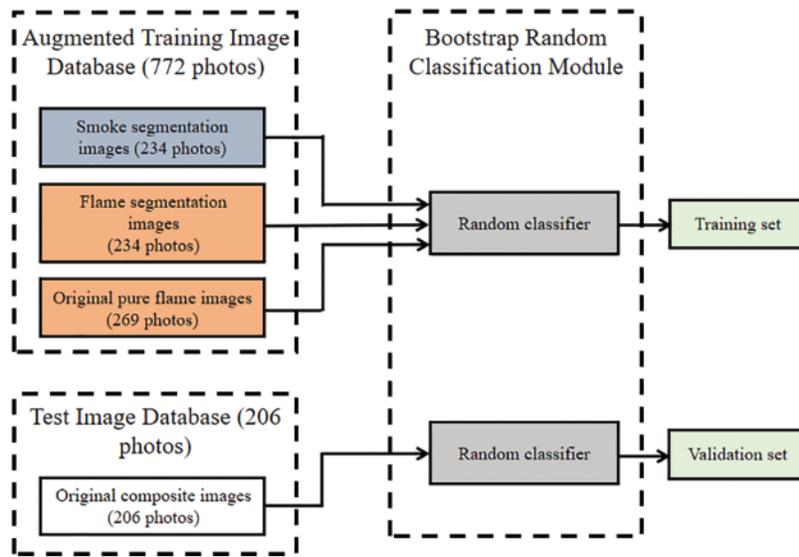
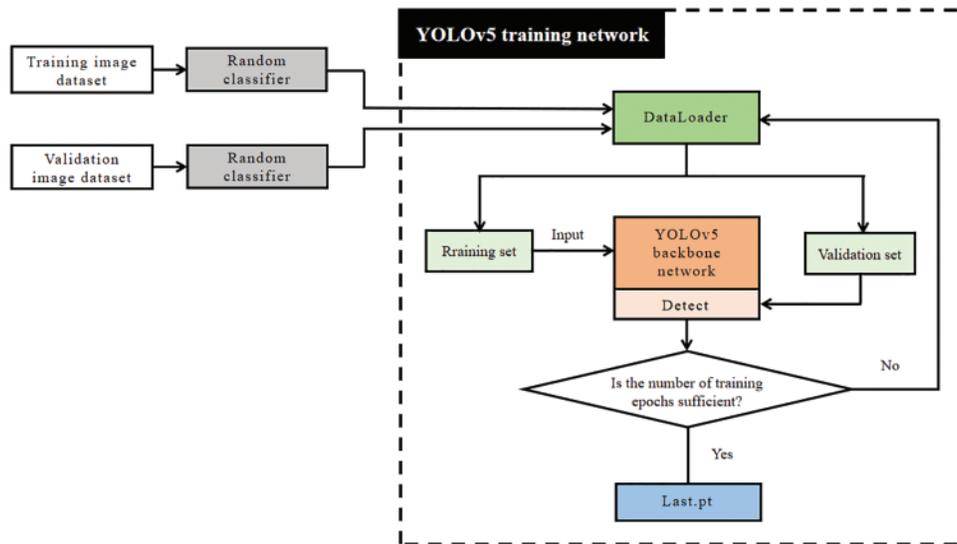**Figure 6:** Schematic diagram of random extraction of training data by bootstrap

Figure 7: Schematic diagram of random cross-validation application

4 Evaluation and Comparison of Experimental Results

4.1 Data Preparation for Small Dataset Training

In this section, by comparing about 2000 images processed by the threshold segmentation method, we find three types of objects that have the most significant impact on the flame recognition work after the threshold segmentation method: sunlight (including sunlight and the burning cloud target caused by sunlight exposure, etc.), reflect (including flame reflected light, light reflected light, the sun reflected light, etc.), and lamp (including screens, various lights, etc.). Finally, we randomly selected 475 images containing positive and negative samples from the original data set. The specific data preparation flow chart is shown in Fig. 8:
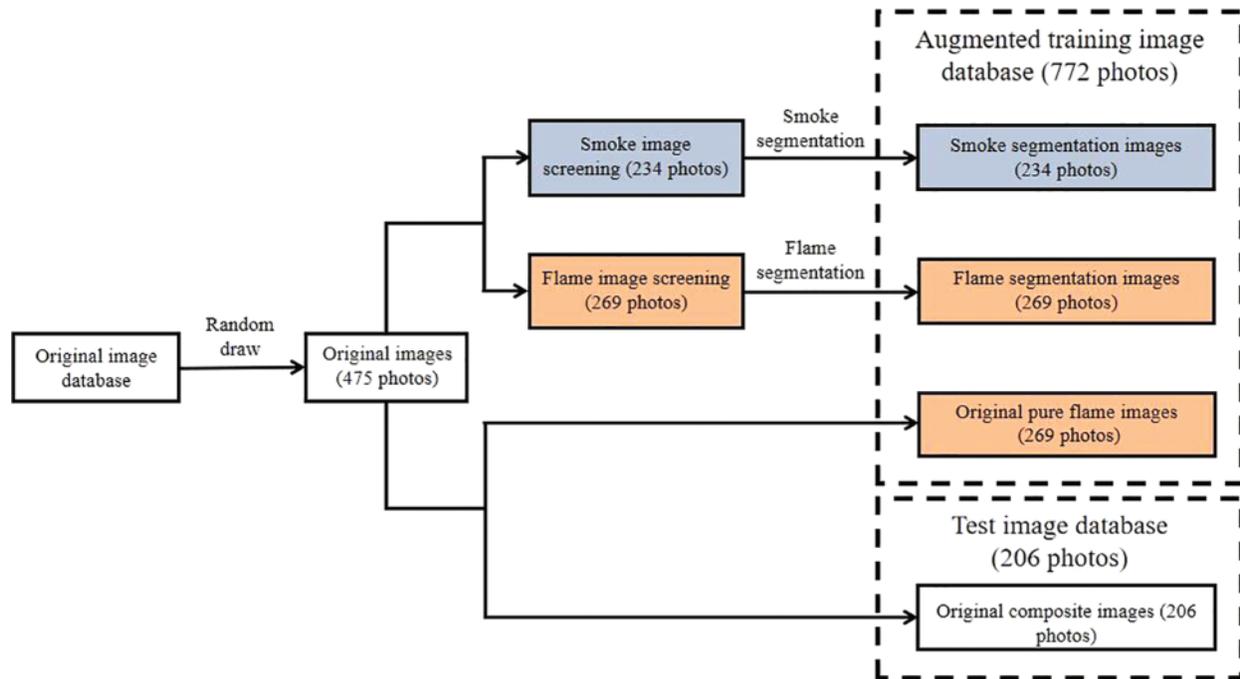
Figure 8: Distribution of training set and validation set

4.2 Evaluation Method for Improved Flame Recognition Model

Determining the prediction results needs to applied the concept of IoU intersection and union ratio. Simply put, it is to calculate the intersection area and union area of the prediction frame and the annotation frame. When the IoU value is larger, it indicates that the prediction result is closer to the annotation result. When the IoU exceeds the set threshold, we judge it as a positive class; otherwise, it is judged as a negative class.

$$IoU = \frac{U_C}{U_A + U_B} \tag{5}$$

The precision rate can reflect the proportion of correctly predicted positive samples in all predicted positive samples. The formula of precision rate is as follows:

$$precision = \frac{TP}{TP + FP} \tag{6}$$

The recall rate is set for the original sample, and it represents how many of the positive samples in the total sample were correctly predicted. The formula of the recall rate is as follows:

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

mAP (Mean Average Precision), which is the mean average precision, is calculated by computing the sum of the average precision of all categories divided by the total number of categories. The average precision of AP (Average Precision) can understand the area enclosed by the curve constructed by the precision rate and the recall rate under the condition of different confidence levels of this class as an

indicator to measure the detection accuracy in the target detection task. The formula for calculating *mAP* is as follows:

$$mAP = \frac{AP_1 + AP_2 + \cdots + AP_n}{n} \tag{8}$$

4.3 Comparison of Improved Flame Recognition Segmentation Methods

The image after the original image is processed by the fusion formulas (2)–(4) is shown in Fig. 9 below.

Figure 9: Improved flame segmentation renderings (a) house fire (b) dark place fire

The image on the left is the original image, the image in the middle is that the processed image and the original image are fused with a weight of 0.5; the image on the right is the binary image after foreground extraction, the white part is the extracted area after segmentation, and the black part is needed after segmentation. The foreground segmentation algorithm can remove the background as much as possible on the basis of preserving the complete flame. It is guaranteed that the flame characteristics are not lost. Compared with the traditional flame segmentation image based on RGB, HSV and YCbCr algorithm, it has a better effect, as shown in Fig. 10 below.

The flame data set processed by the algorithm can be more effectively put into the neural network for training, making it easier for the network to learn the image features to achieve a better prediction effect.

4.4 Experimental Results of Improved Yolov5 Network Model

Use the improved flame segmentation method to preprocess the data set images input to the YOLOv5 network and construct the data set according to the previous results. To verify the effectiveness of the random cross-validation module, a total of 7 groups of experiments were set up in this paper. After multiple rounds of experimental comparison, the experimental data were integrated into the following Table 1:
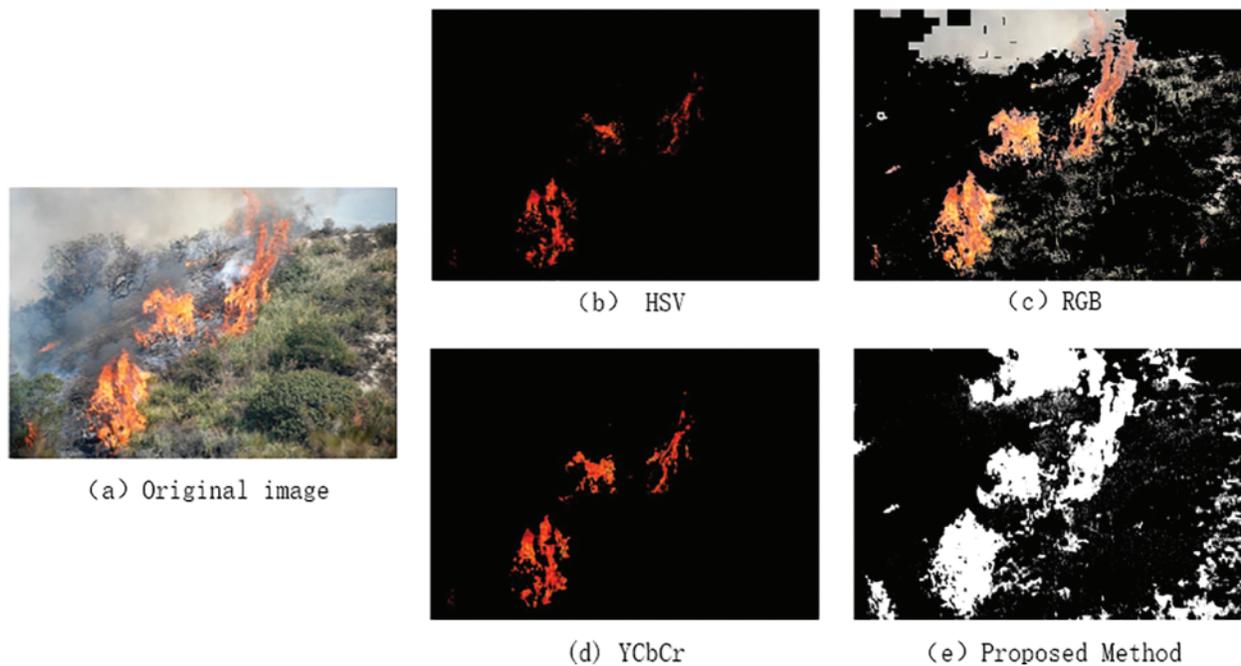
Figure 10: Comparison of the effects of the improved flame segmentation algorithm and the classic flame segmentation algorithm in different color spaces

Table 1: Effective comparative experiment information table

	Description of training data	Validate data description	Random cross-validation	P (fire/all)	R (fire/all)	mAP-0.5 (fire/all)	mAP-0.95 (fire/all)	Learning rate	Training rounds
1	The original image	Original image		0.44/0.17	0.65/0.44	0.41/0.14	0.26/0.09	0.01	300
2	Flame segmentation image	Original image		0.48/0.19	0.68/0.50	0.42/0.15	0.32/0.11	0.0016	800
3	The original image	Flame segmentation image		0.50/0.13	0.78/0.43	0.49/0.12	0.32/0.08	0.0025	300
4	Flame segmentation image	Original image	Yes	0.92/0.86	0.73/0.69	0.86/0.75	0.57/0.51	0.001	300
5	Flame segmentation image and part of original image	Part of original image	Yes	0.90/0.87	0.89/0.82	0.92/0.85	0.68/0.61	0.001	300
6	Flame segmentation image and part of original image and smoke segmented image	Part of original image		0.91/0.88	0.91/0.80	0.92/0.85	0.71/0.62	0.001	200

(Continued)

Table 1: Continued

	Description of training data	Validate data description	Random cross-validation	P (fire/all)	R (fire/all)	mAP-0.5 (fire/all)	mAP-0.95 (fire/all)	Learning rate	Training rounds
7	Flame segmentation image and part of original image+smoke segmented image	Part of original image	Yes	0.96/0.89	0.97/0.84	0.98/0.86	0.74/0.58	0.001	300

In Table 1, *mAP* is the mean average precision, the "mAP-0.5" column is the mean average precision value obtained by each group of experiments when the confidence level is 0.5, and the "mAP-0.95" column is the mean average precision value received by each group of experiments when the confidence level is 0.95. In the results of the four columns "P", "R", "mAP-0.5" and "mAP-0.95", where there are 2 value, the first value represents the recognition effect of the network model on the single-type target of flame, and the second value is the comprehensive recognition effect of the network model for the five categories of "fire", "smoke", "sunshine", "lamp" and "reflect". The learning rate parameter of neural network training needs to be adjusted according to the size of the training set and is determined according to formula (9):

$$lr = \frac{1}{data\_sum} \ (lr \geq 0.001) \tag{9}$$

The data_sum parameter is the total number of images in the dataset. Due to the small size of the data, the model trained based on the original 600 or so labeled flame image datasets, that is, the detection model obtained from the first group of experimental training, has not performed well in various parameters of the detection effect. The model has a preliminary recognition effect on each category of image in the validation set; however, the recall rate is low, the missed detection rate is high, the accuracy rate is low, and the misjudgment rate is high.

The flame segmentation images processed by the improved threshold segmentation method based on RGB color space are added to the data set, and experiments 2 and 3 are designed for analysis. Compared the network parameter maps in the training process of experiment 2 and experiment 3, the segmented flame images were used as the test set. The detection model obtained better detection effect: the accuracy P of flame recognition was improved from 0.48 to 0.5; the R was increased from 0.68 to 0.78. However, the growth stability of evaluation parameters of the model is poor, and it is more difficult for the detection model to select a perfect flame range in the segmented image in the test image.

The fourth group of experiments added a random cross-validation module on the basis of the second group of experiments. Before each round of training in the network training process, the flame segmentation image training data set and the original image test set was randomly divided (the images of the training set and test set do not overlap). The experimental results of combined experiment 4 and experiment 2 are analyzed and compared. After adding the random test module, the accuracy P of flame identification was increased from 0.48 to 0.9, and the IoU was increased from 0.42 to 0.86; The P of all categories was increased from 0.19 to 0.86, and the average IoU was increased from 0.15 to 0.75. So the effect of the network on flame recognition is further improved, the target recognition accuracy rate is higher, and the recall rate is higher. But during the training process, since the training data and test data are constantly changing, the training results will also be continuously adjusted, and its parameters will fluctuate accordingly.

In order to solve the problem of large fluctuations in the growth of PR parameters during the cross-validation process, on the basis of the fourth group of experiments, the original data set was further adjusted. 206 images with obvious flame characteristics and few negative samples in the original image were selected and added to the training data set to expand the training set, and the fifth group of experiments was designed to carry out. Combining Experiment 5, Experiment 4, and Experiment 2 can be seen from the analysis and comparison of Table 1. After the data set is divided into a new one, the IoU of the network model for flame identification was increased from 0.86 to 0.92; The average IoU of all categories was increased from 0.75 to 0.85. In the test process of the verification set of Experiment 5, the network model can already well grasp the characteristics of the flame; the recognition effect of the flame is good and very stable, and the confidence of the recognition target is also high. In experiment 5, the comparison of the recognition effect and the comparison of the parameter changes are shown in Fig. 11:

Figure 11: Network recognition results of experiment 5

After adding random cross-validation, the network model significantly improved the recognition effect of flames. Still, the recognition effect of smoke in all categories is not significantly improved. The PR parameter map of the best training model in the previous five sets of experiments is used as a reference, as shown in Fig. 12a:

As shown in the PR diagram of all network categories, the area enclosed by the smoke, lamp, and coordinate axis is less than the total area of "all classes" and the prediction effect is poor. Although these two categories are not directly related to the flame recognition work, improving the recognition ability of the "lamp" category can improve the network model's ability to distinguish between the "fire" category and the "lamp" category and reduce the misjudgment of flame identification. The "smoke" category is also an essential criterion in the fire scene, and it is strongly related to the generation of flames in the background. Therefore, improving the detection ability of the "smoke" category is also of great significance in the work of flame recognition.
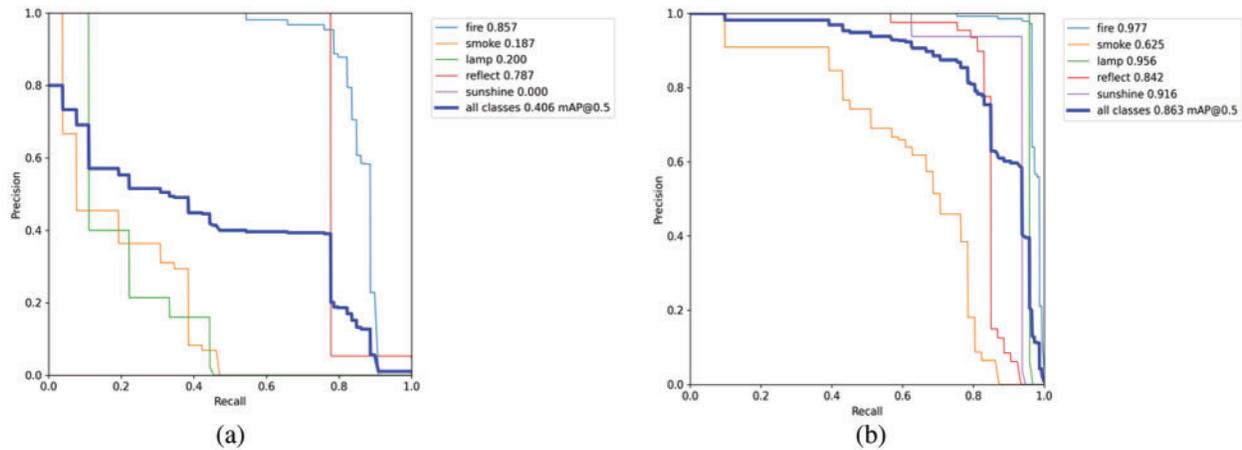
Figure 12: PR chart of each category (a) experiment 5 (b) experiment 6

In order to enhance the effect of flame recognition, we selected 234 original images with prominent smoke characteristics from the source data set for smoke segmentation. As a data set to enhance the effect of smoke recognition, we divided them into the training set and set up the sixth group of experiments. So far, the network training set contains 269 original images, 269 flame images, 234 flame segmentation images, and smoke segmentation images, for a total of 772 images in the training set. The test set images contain a total of 206 comprehensive images of all categories of test cases. The sixth training group is performed on the newly divided data set (Fig. 12b).

As shown in Fig. 12b above, the detection model of the sixth group of experiments has a good improvement in the detection ability of "smoke" and "lamp", and the recognition effect of other categories has also been improved. In the seventh group of experiments, the random cross-validation module was added again on the basis of the sixth group of experiments, and a better recognition effect was obtained, shown as Fig. 13. The recognition accuracy of the smoke part was improved, and the related evaluation indicators for all categories were improved.

The experimental results in Table 2, show that in the network training divided by the bootstrap, the training evaluation parameters will fluctuate significantly in the early stage due to the frequent changes in the data set. After multiple training rounds, when the network masters the model features, the evaluation parameters will tend toward stability. After adding the bootstrap, the network recognition effect has been further improved, and the recognition stability and accuracy have improved, breaking through the bottleneck that the recognition effect of small training sets is difficult to improve.

It is can be seen from Table 2 that after adding the bootstrap, the recognition accuracy of flame has reached 0.96, the recall rate has reached 0.966, the handover merger ratio has reached 0.977, the recognition accuracy rate of all categories reached 0.889, the recall rate reached 0.844, and the handover merger ratio reached 0.863. The experimental results show that the network recognition effect has been further improved, and the recognition stability and accuracy have improved.
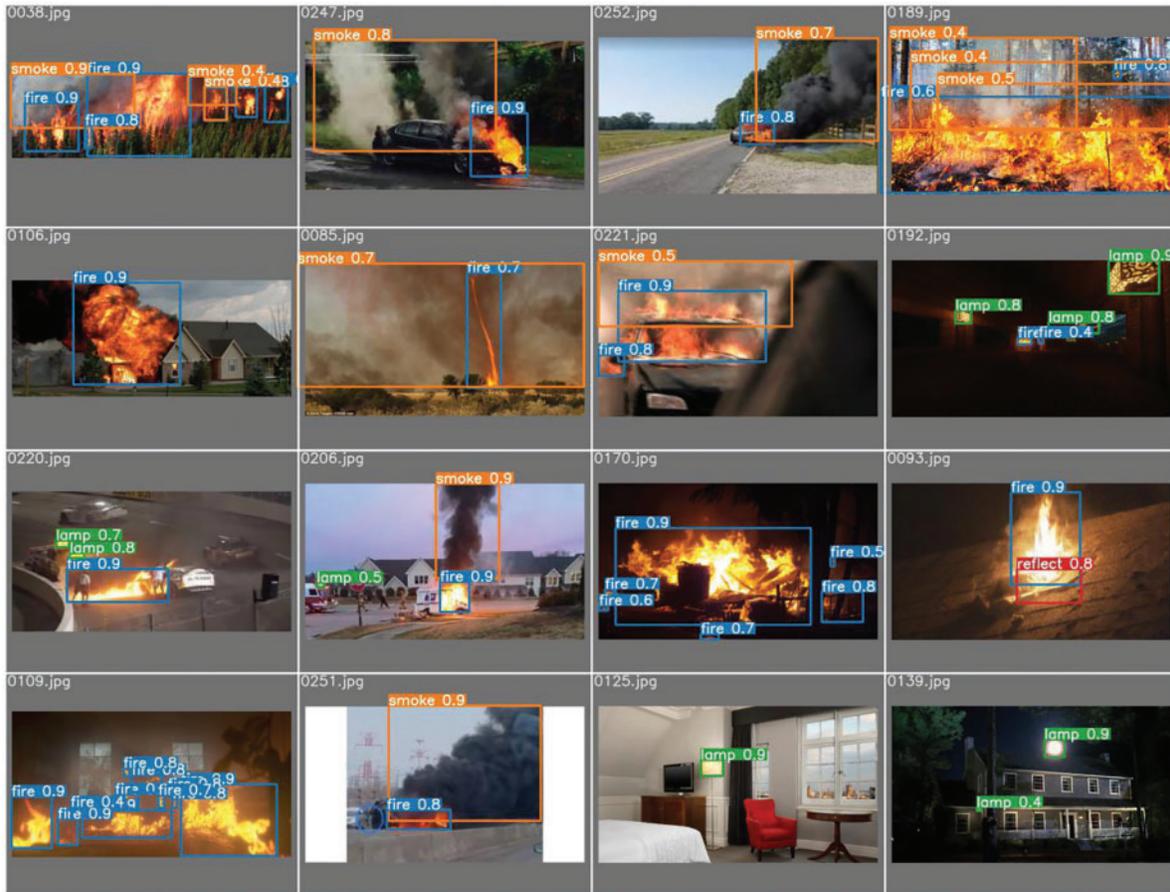
Figure 13: Network recognition results of experiment 7

Table 2: Parameter of recognition effect of each category in the improved classification network

	Classification name	Category description	Precision	Recall	mAP-0.5	mAP-0.95
1	Fire	Open fire, dark fire	0.96	0.966	0.977	0.741
2	Smoke	Solid small particle smoke due to combustion	0.68	0.569	0.625	0.325
3	Lamp	Light	0.934	0.958	0.956	0.629
4	Reflect	Reflected light from light sources such as flames, sunlight, lights, etc.	0.935	0.811	0.842	0.548
5	Sunshine	Sunlight and sunlit fire clouds and other rays of light	0.936	0.917	0.916	0.636
6	All	All categories above	0.889	0.844	0.863	0.576

5 Conclusion

This paper mainly improves the threshold segmentation method and the YOLOv5 neural network training strategy in the flame detection work. The ratio of different band thresholds is used to replace the threshold in the traditional flame threshold segmentation algorithm for flame segmentation, and the flame images preprocessed by the method are trained by YOLO neural network. The bootstrapping and cross-validation are introduced into the YOLOv5 network model. The experimental results show that the proposed method has a small dependence on the dataset and can detect the flame features more and more quickly and accurately than before improvement. However, the algorithm still has the problem of insufficient detection accuracy for small target flame. The root cause of this problem is that in the convolution neural network, the algorithm often has to perform multiple downsampling operations, which leads to the lack of information of the small target object in the deep feature extraction layer. The algorithm thus lacks accuracy and stability for small target object detection. In the future, detection methods suitable for smaller-scale training data sets will be explored to further reduce the time cost of model training and improve the accuracy of model detection. In later experiments, we will consider adding attention mechanism module to improve the algorithm.

Author Contributions: Zhao Yuchen, Wu Shulei, Wang Yaoru, Chen Huandong, Zhang xianyao and Zhao Hongwei contributed equally to method design, experimental analysis, and manuscript writing. Wu Shulei and Chen Huandong were responsible for data provision and funding acquisition. All authors reviewed the final version of the manuscript and consented to publication.

Availability of Data and Materials: The data reported in this study are included in the article.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1]   A. Rostami, R. Shah-Hosseini, S. Asgari, A. Zarei, M. Aghdami-Nia *et al.,* "Active fire detection from Landsat-8 imagery using deep multiple kernel learning," *Remote Sensing*, vol. 14, no. 4, pp. 992–1016, 2022.

[2]   U. A. Bhatti, M. X. Huang, D. Wu, Y. Zhang, A. Mehmood *et al.,* "Recommendation system using feature extraction and pattern recognition in clinical care systems," *Enterprise Information Systems*, vol. 13, no. 3, pp. 329–351, 2019.

[3]   B. Kim and J. Lee, "A video-based fire detection using deep learning models," *Applied Sciences*, vol. 9, no. 14, pp. 2862–2881, 2019.

[4]   A. Costeaand and P. Schiopu, "New design and improved performance for smoke detector," in *2018 10th Int. Conf. on Electronics, Computers and Artificial Intelligence (ECAI)*, Iasi, Romania, pp. 1–7, 2018.

[5]   K. -T. Hsu, B. -S. Wang, C. -H. Lin, B. -Y. Wang and W. -P. Chen, "Development of mountable infrared-thermal image detector for safety hamlet," in *2021 Int. Conf. on Electronic Communications, Internet of Things and Big Data(ICEIB)*, Yilan County, Taiwan, pp. 105–109, 2021.

[6]   Ş. Topuk and E. Tosun, "Design of the flame detector filter vith wavelet transform and fuzzy logic algorithm," in *2020 12th Int. Conf. on Electrical and Electronics Engineering (ELECO)*, Bursa, Turkey, pp. 207–211, 2020.

[7]   P. Barmpoutis, T. Stathaki, K. Dimitropoulos and N. Grammalidis, "Early fire detection based on aerial 360-degree sensors, deep convolution neural networks and exploitation of fire dynamic textures," *Remote Sensing*, vol. 12, no. 19, pp. 3177–3194, 2020.

[8]   H. Alqourabah, A. Muneer and S. M. Fati, "A smart fire detection system using IoT technology with automatic water sprinkler," *International Journal of Electrical & Computer Engineering*, vol. 11, no. 4, pp. 2088–8708, 2021.

[9]   V. Kodur, P. Kumar and M. M. Rafi, "Fire hazard in buildings: Review, assessment and strategies for improving fire safety," *PSU Research Review*, vol. 4, no. 1, pp. 1–23, 2020.

[10]   M. Aliff, N. S. Sani, M. I. Yusof and A. Zainal, "Development of fire fighting robot (QROB)," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 1, 2019.

[11]   X. Ren, C. Li, X. Ma, F. Chen, H. Wang *et al.,* "Design of multi-information fusion based intelligent electrical fire detection system for green buildings," *Sustainability*, vol. 13, no. 6, pp. 3060–3405, 2021.

[12]   S. Kaewunruen and N. Xu, "Digital twin for sustainability evaluation of railway station buildings," *Frontiers in Built Environment*, vol. 4, pp. 1–10, 2018.

[13]   Y. Z. Li and H. Ingason, "Overview of research on fire safety in underground road and railway tunnels," *Tunnelling and Underground Space Technology*, vol. 81, no. NOV, pp. 568–589, 2018.

[14]   F. Mirahadi, B. McCabe and A. Shahi, "IFC-centric performance-based evaluation of building evacuations using fire dynamics simulation and agent-based modeling," *Automation in Construction*, vol. 101, no. MAY, pp. 1–16, 2019.

[15]   V. Jaiswal, V. Sharma and S. Varma, "MMFO: Modified moth flame optimization algorithm for region based RGB color image segmentation," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 1, pp. 196–201, 2020.

[16]   S. Lenka, B. Vidyarthi, N. Sequeira and U. Verma, "Texture aware unsupervised segmentation for assessment of flood severity in UAV aerial images," in *IGARSS 2022–2022 IEEE Int. Geoscience and Remote Sensing Symp.,* Kuala Lumpur, Malaysia, pp. 7815–7818, 2022.

[17]   W. Haifeng and Z. Yi, "A fast one-dimensional Otsu segmentation algorithm based on least square method," *Journal of Advanced Oxidation Technologies*, vol. 21, no. 2, 2018.

[18]   M. Seo, T. Goh, M. Park and S. W. Kim, "Detection of internal short circuit for lithium-ion battery using convolutional neural networks with data pre-processing," *International Journal of Electronics and Electrical Engineering*, vol. 10, no. 1, pp. 76–89, 2017.

[19]   W. Lang, Y. Hu, C. Gong, X. Zhang, H. Xu *et al.,* "Artificial intelligence-based technique for fault detection and diagnosis of EV motors: A review," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 1, pp. 384–406, 2021.

[20]   S. Wilson, S. P. Varghese, G. A. Nikhil, I. Manolekshmi and P. G. Raji, "A comprehensive study on fire detection," in *2018 Conf. on Emerging Devices and Smart Systems (ICEDSS)*, Tiruchengode, India, pp. 242–246, 2018.

[21]   J. Samuel, A. Wetzel, S. Chapman, E. Tollerud, P. F. Hopkins *et al.,* "Planes of satellites around Milky Way/M31-mass galaxies in the FIRE simulations and comparisons with the local group," *Monthly Notices of the Royal Astronomical Society*, vol. 504, no. 1, pp. 1379–1397, 2021.

[22]   S. D. Thepade, J. H. Dewan, D. Pritam and R. Chaturvedi, "Fire detection system using color and flickering behaviour of fire with Kekre's luv color space," in *2018 Fourth Int. Conf. on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, pp. 1–6, 2018.

[23]   F. Gong, C. Li, W. Gong, X. Li, X. Yuan *et al.,* "A real-time fire detection method from video with multifeature fusion," *Computational Intelligence and Neuroscience*, vol. 2019, no. 3, pp. 1–17, 2019.

[24]   B. Cai, Y. Wang, J. Wu, M. Wang, F. Li *et al.,* "An effective method for camera calibration in defocus scene with circular gratings," *Optics and Lasers in Engineering*, vol. 114, no. MAR, pp. 44–49, 2019.

[25] G. Shi, T. Huang, W. Dong, J. Wu and X. Xie, "Robust foreground estimation via structured Gaussian scale mixture modeling," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 4810–4824, 2018.

[26] U. A. Bhatti, Z. Yu, J. Chanussot, Z. Zeeshan, L. Yuan *et al.,* "Local similarity-based spatial-spectral fusion hyperspectral image classification with deep CNN and Gabor filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2021.

[27] U. A. Bhatti, Z. Zeeshan, M. M. Nizamani, S. Bazai, Z. Yu *et al.,* "Assessing the change of ambient air quality patterns in Jiangsu Province of China pre-to post-COVID-19," *Chemosphere*, vol. 288, no. 2, pp. 132569–132579, 2022.

[28] S. Wu, F. Zhang, H. Chen and Y. Zhang, "Semantic understanding based on multi-feature kernel sparse representation and decision rules for mangrove growth," *Information Processing and Management*, vol. 59, no. 2, pp. 102813–102828, 2022.

[29] S. Wu, H. Chen, Y. Bai and G. Zhu, "A remote sensing image classification method based on sparse representation," *Multimedia Tools and Applications*, vol. 75, no. 19, pp. 12137–12154, 2016.

[30] C. Wang, Q. Luo, X. Chen, B. Yi and H. Wang, "Citrus recognition based on YOLOv4 neural network," *Journal of Physics: Conference Series*, vol. 1820, no. 1, pp. 012163–012172, 2021.

[31] Y. Zhao, Y. Shi and Z. Wang, "The improved YOLOV5 algorithm and its application in small target detection," in *Int. Conf. on Intelligent Robotics and Applications*, Harbin, China, vol. 13458, pp. 679–688, 2022.

[32] Y. Xiao, A. Chang, Y. Wang, Y. Huang, J. Yu *et al.,* "Real-time object detection for substation security early-warning with deep neural network based on YOLO-V5," in *2022 IEEE IAS Global Conf. on Emerging Technologies (GlobConET)*, Arad, Romania, pp. 45–50, 2022.

[33] K. Muhammad, J. Ahmad, I. Mehmood, S. Rho and S. W. Baik, "Convolutional neural networks based fire detection in surveillance videos," *IEEE Access*, vol. 6, pp. 18174–18183, 2018.

[34] C. Chaoxia, W. Shang and F. Zhang, "Information-guided flame detection based on faster R-CNN," *IEEE Access*, vol. 8, no. 8, pp. 58923–58932, 2020.

[35] Y. Chen, Y. Zhang, J. Xin, G. Wang, L. Mu *et al.,* "UAV image-based forest fire detection approach using convolutional neural network," in *2019 14th IEEE Conf. on Industrial Electronics and Applications (ICIEA)*, Xi'an, China, pp. 2118–2123, 2019.

[36] P. Li and W. Zhao, "Image fire detection algorithms based on convolutional neural networks," *Case Studies in Thermal Engineering*, vol. 19, pp. 100625–100636, 2020.

[37] U. A. Bhatti, Z. Ming-Quan, H. Qing-Song, S. Ali, A. Hussain *et al.,* "Advanced color edge detection using clifford algebra in satellite images," *IEEE Photonics Journal*, vol. 13, no. 2, pp. 1–20, 2021.

[38] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 779–788, 2016.

[39] D. Li, Y. W. Chen, J. Li, L. Cao, U. A. Bhatti *et al.,* "Robust watermarking algorithm for medical Images based on accelerated-KAZE discrete cosine transform," *IET Biometrics*, vol. 11, no. 6, pp. 534–546, 2022.

[40] Q. Zhang, J. Xu, L. Xu and H. Guo, "Deep convolutional neural networks for forest fire detection," in *2016 Int. Forum on Management, Education and Information Technology Application*, Changchun, China, pp. 568–575, 2016.

[41] N. S. Bakri, R. Adnan, A. M. Samad and F. A. Ruslan, "A methodology for fire detection using colour pixel classification," in *2018 IEEE 14th Int. Colloquium on Signal Processing & Its Applications (CSPA)*, Penang, Malaysia, pp. 94–98, 2018.

[42] L. Yu and C. Zhou, "Determining the best clustering number of K-means based on bootstrap sampling," in *2nd Int. Conf. on Data Science and Business Analytics (ICDSBA)*, Changsha, China, pp. 78–83, 2018.

[43] B. Mu, T. Chen and L. Ljung, "Asymptotic properties of hyperparameter estimators by using cross-validations for regularized system identification," in *2018 IEEE Conf. on Decision and Control (CDC)*, Miami, FL, USA, pp. 644–649, 2018.

[44] P. Ju, X. Lin and N. B. Shroff, "Understanding the generalization power of overfitted NTK models: 3-layer *vs.* 2-layer (Extended Abstract)," in *2022 58th Annual Allerton Conf. on Communication, Control, and Computing (Allerton)*, Monticello, IL, USA, pp. 1–2, 2022.