ORIGINAL RESEARCH
published: 08 March 2022
doi: 10.3389/fenrg.2022.848754

Multi-Scale Video Flame Detection for Early Fire Warning Based on Deep Learning

Peiwen Dai[1], Qixing Zhang[1]*, Gaohua Lin[1], Muhammad Masoom Shafique[1], Yinuo Huo[1], Ran Tu[2] and Yongming Zhang[1]

[1]State Key Laboratory of Fire Science, University of Science and Technology of China, Hefei, China, [2]College of Mechanical Engineering and Automation, Huaqiao University, Xiamen, China

The widespread use of renewable energy resources requires more immediate and effective fire alarms as a preventive measure. The fire is usually weak in the initial stages, which is not conducive to detection and identification. This paper validates a solution to resolve that problem by a flame detection algorithm that is more sensitive to small flames. Based on Yolov3, the parallel convolution structure of Inception is used to obtain multi-size image information. In addition, the receptive field of the convolution kernel is increased with the dilated convolution so that each convolution output contains a range of information to avoid information omission of tiny flames. The model accuracy has improved by introducing a Feature Pyramid Network in the feature extraction stage that has enhanced the feature fusion capability of the model. At the same time, a flame detection database for early fire has been established, which contains more than 30 fire scenarios and is suitable for flame detection under various challenging scenes. Experiments validate the proposed method not only improves the performance of the original algorithm but are also advantageous in comparison with other state-of-the-art object detection networks, and its false positives rate reaches 1.2% in the test set.

Keywords: flame detection, video fire detection, multi-scale, deep learning, early fire

OPEN ACCESS

Edited by:
Weiguang An,
China University of Mining and Technology, China

Reviewed by:
Xueming Shu,
Tsinghua University, China
Chenqiang Gao,
Chongqing University of Posts and Telecommunications, China

****Correspondence:***
Qixing Zhang
qixing@ustc.edu.cn

Specialty section:
This article was submitted to Sustainable Energy Systems and Policies,
a section of the journal Frontiers in Energy Research

Received: *05 January 2022*
Accepted: *15 February 2022*
Published: *08 March 2022*

Citation:
Dai P, Zhang Q, Lin G, Shafique MM, Huo Y, Tu R and Zhang Y (2022) Multi-Scale Video Flame Detection for Early Fire Warning Based on Deep Learning. Front. Energy Res. 10:848754. doi: 10.3389/fenrg.2022.848754

INTRODUCTION

Renewable energy sources is playing an increasingly important role in industry (Qazi et al., 2019). Therefore, its security problem is attracting widespread attention. We can see numerous studies on Hydrogen safety, lithium-ion battery safety, and Photovoltaic safety (Yang et al., 2018; Ould Ely et al., 2019; Abohamzeh et al., 2021; Fang et al., 2021). However, few studies are reported in the literature for efficient flame detection in the case of fire accidents for renewable energy sources. Because of the unique characteristics of renewable energy resources, their fire situation is complicated, and the immediacy in fire detection and accuracy of fire alarms is necessary for reducing fire hazards. Traditional fire detection technologies detect fire according to the characteristic signals of fire, such as temperature, combustion gas, aerosol, etc. (Xu, 2020). However, such characteristic signals are weakened gradually in the process of propagation, therefore the traditional contact detector will be restricted by the height and area of the detection space. With the development of digital image processing, video fire detection technology has been proposed and researched. Video fire detection technology that does not depend on contact characteristic parameters became more advantageous in the fire detection domain due to its advantages like a fast response, visualization, and broader detection space. So, vision-based fire detection systems can play a decisive role in the flame detection of renewable energy sources.

The traditional video fire detection method is based on the characteristics of the flame. The static characteristics of the flame include color, shape, number of sharp angles, and circularity, while the dynamic characteristics include flicker frequency and flame area change rate. Yamagishi et al. (Yamagishi and Yamaguchi, 1999) innovatively processed the HSV color space and extracted the flame area by taking advantage of the changing characteristics of the color and saturation in the flame area. Izquierdo and Borges (Borges and Izquierdo, 2010) realized fire detection by Bayes classifier using changes in the shape, boundary, and area of the flame region and other additional features. Dimitropoulos K and Barmpoutis (Dimitropoulos et al., 2014) proposed a fire detection method based on multi-feature extraction, which simultaneously established a fire model based on flame scintilla feature, dynamic texture feature, color feature, and spatiotemporal energy. However, the traditional video flame detection method based on flame features has its limitations. The algorithm mostly uses static images, lacks dynamic feature extraction, and is susceptible to interference from the shadow, brightness, energy, and other factors. The false positive rate is high, and the detection sensitivity is overly dependent on the algorithm parameters.

Since the rise of deep learning in 2012 (Ghali et al., 2020), it has made outstanding achievements in image classification and object detection, causing a new upsurge in the fields of artificial intelligence and computer vision. Among them, the convolutional neural network is the most outstanding one in image data processing. Convolutional neural network (CNN) is a kind of feedforward neural network with a deep structure (Lecun et al., 1998), which includes convolution computation. And it is a research hotspot in the field of semantic analysis and image recognition. CNN has a weight sharing network structure similar to a biological neural network, which reduces the complexity of the network model by reducing the number of weights, which not only reduces the training parameters but also greatly improves the training speed.

As a branch of computer vision, video fire detection also begins to introduce deep learning. Frizzi (Frizzi et al., 2016) uses a 9-layer convolutional neural network to extract features from images and realizes the classification of smoke and flame through sliding window search, which is very fast. Compared with traditional video fire detection methods, this method has better classification performance, indicating that it is promising to use CNN to detect fire in video. Yong-Jin Kim (Young-Jin and Eun-Gyung, 2017)tries to apply Faster RCNN to flame detection, and Shen (Shen et al., 2018)simplifies the Yolo (You Only Look Once) network to carry out flame detection, both of which achieve good results, indicating that the flame detection method based on deep learning is superior to the traditional video fire detection method in performance.

The early stage of fire is the best stage to extinguish the fire, so the fire detection and alarm at this stage are particularly important. However, the early flame of fire is weak, so it is easy to be ignored by the detection model. To solve this problem, this paper proposes a fire detection and identification method based on improved Yolov3. Yolov3 (You only look once v3) is an excellent object detector with good performance in both aspects of accuracy and speed. Based on this, we hope to improve its ability to identify small objects and introduce multi-scale convolution and dilated convolution into the backbone network to improve its ability to identify flames at different scales. At the same time, the idea of FPN (Feature Pyramid Networks) is used to improve the feature extraction network of Yolov3. The proposed method strengthens the feature fusion and reuses high-level features to achieve the purpose of improving accuracy. In addition, this paper has established a flame database for early fires, which involves a variety of fire scenarios to establish a foundation for future flame detection research.

RELATED WORK

There are many applications of deep learning methods in flame detection. Some studies try to combine the traditional video flame detection method with deep learning, and first carry out feature extraction and then use convolutional neural network for recognition. Chen et al. (Zhong et al., 2020) designed a flame detection method based on multi-channel convolutional neural network, the OTSU algorithm was used to extract the flame color contour and dynamic features, and then the three features were input into the three-channel convolutional neural network for detection and recognition. Compared with traditional methods, the accuracy is improved, but the method of training specific features has some problems of over-fitting. Otabek Khudayberdiev et al. (Khudayberdiev and Butt, 2020). combined PCA(Principal component analysis) and CNN, extracted data features using PCA, and CNN conducted inspection and classification. MobileNet was selected as the backbone to simplify the size of the model but there is a lack of accuracy.

Some researchers choose to carry out transfer learning, which is to apply the pre-trained deep CNN architecture for the development of fire detection systems. Mohit et al. (Dua et al., 2020) believed that the traditional use of the CNN method to carry out flame detection using balanced data sets was not in line with the actual situation, so they proposed to use unbalanced data sets with more non-fire pictures. They used two models, VGG (Visual Geometry Group) and MobileNet, for flame detection, and the experimental results were superior to the traditional CNN method. Jivitesh (Sharma et al., 2017) also used unbalanced data sets for training detection, and in his experiment, Resnet50 outperformed VGG16.

Some researchers believe that the detection of static frames has its limitations, so the deep learning method is considered to identify the dynamic characteristics of flames in the video. Lin et al. (Lin et al., 2019) proposed a joint detection framework based on Faster RCNN (Faster Regions with CNN features) (Ren et al., 2016) and 3D CNN (Tran et al., 2015), where RCNN is mainly used to select the suspected fire area for preliminary identification, while 3D CNN is used to extract temporal information and combine the static features and temporal features of smoke. Kim et al. (Kim and Lee, 2019)first used

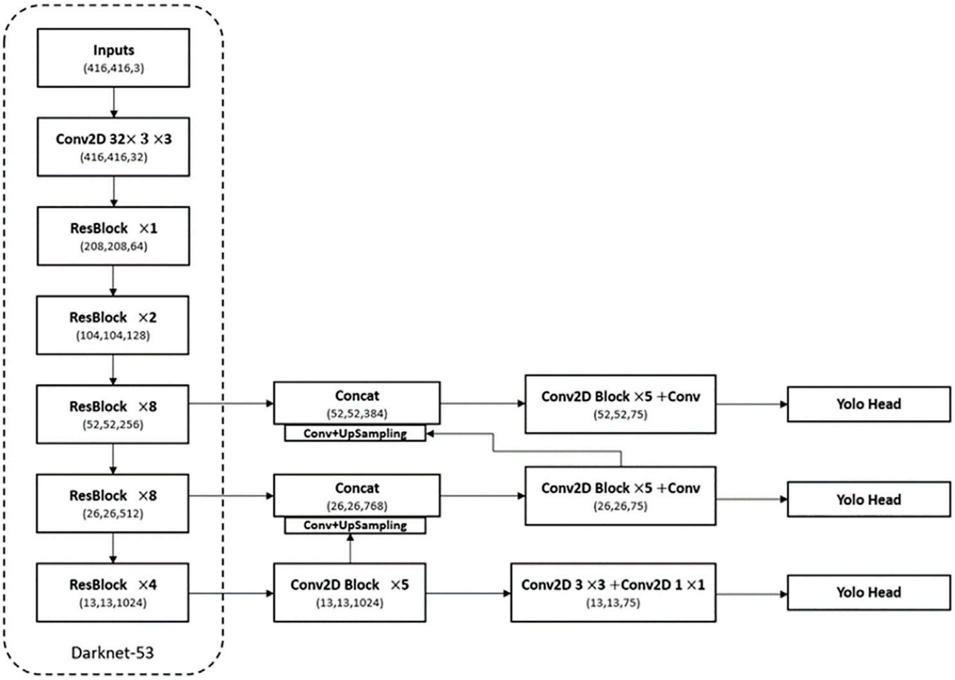


FIGURE 1 | Structure of Yolov3 which includes backbone and feature extraction structure.

Faster RCNN to detect the suspected fire area, and then used LSTM (Long Short-Term Memory) to judge whether there was a flame from the space-time characteristics. Although this kind of method improves the accuracy of fire detection compared with the image-based method, the huge structure of the model is limited in practical application.

Fire detection based on deep learning mainly revolves around detection accuracy and model size. At present, most of the data sets in the literature are flame images with clear texture and large size, but there is a small proportion of research being carried out for early flame detection that is mandatory to meet the real world applications. Therefore, based on the characteristics of early flames, this paper proposes a detection method for small flames, and at the same time strikes a balance between accuracy and model size. Pu Li (Li and Zhao, 2020) et al. summarized the current advanced object detection algorithm and selected four representative models, such as Faster-RCNN, R-FCN (Dai et al., 2016), SSD, and Yolov3, to test the fire data set. The results showed that Yolov3 had the best performance in flame detection. Therefore, this paper considers Yolov3 aiming to research small flame detections.

THE PROPOSED METHOD

Object detection is a common method used in fire detection by computer vision technology. Yolov3 (Redmon and Farhadi, 2018) is an excellent object detection network with a balance between speed and accuracy. It has three times the detection speed while achieving the same accuracy as SSD (Liu et al., 2016). Many experiments have shown that Yolov3 (You only look once v3) is the state-of-the-art object detector with good performance in both aspects of accuracy and speed (Zhang et al., 2020). Based on the results of Puli's study (Li and Zhao, 2020), we chose Yolov3 to improve on early flame detection.

The overall structure of the Yolov3 algorithm is shown in **Figure 1**, which can be divided into three parts, including backbone, multi-scale feature extraction structure, and the output. Our model uses parallel convolution structure to get semantic information of different sizes and uses dilated convolution to increase the reception field. Feature Pyramid Networks is used in feature extraction structure to strengthen the utilization of information of different feature layers. Our model has a total of 45,774,941 parameters and a size of 174 MB, and the overall structure of the network is shown in **Figure 2**.

Backbone

In **Figure 1** we can see that there are many residual modules in the backbone network Darknet53 of Yolov3. The structure of the residual module is shown in **Figure 3**. In the residual module, a convolution calculation with the size of 3 × 3 and an activation function processing is first carried out, and then the layer is temporarily saved. Then, this layer is convoluted twice with sizes of 1 × 1, 3 × 3 respectively. Finally, this convolutional layer is merged with the previously saved convolutional layer by jumping connection and output. It can be found that the convolution scale and convolution method in the backbone network of Yolov3 are relatively single. As a result, Yolov3 is slightly weak in multi-scale recognition. Therefore, the residual module should be

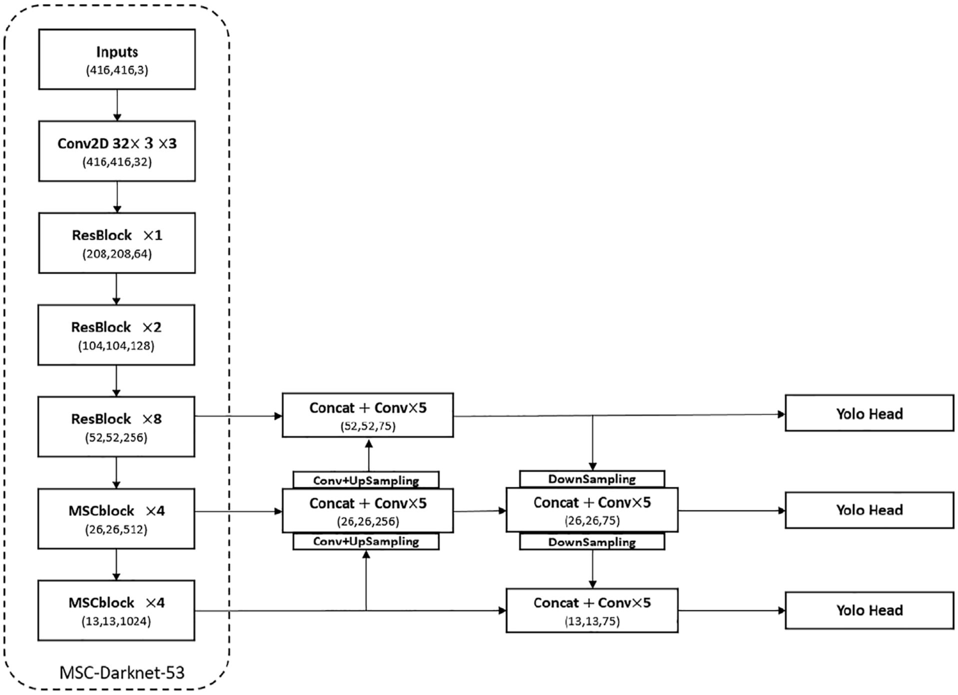


FIGURE 2 | Structure of the proposed method which includes backbone and feature extraction structure.

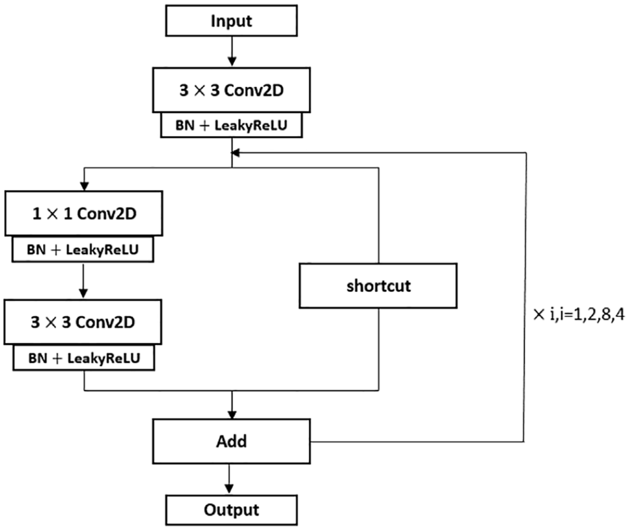


FIGURE 3 | Cyclic structure of residual module.

improved from these two directions, and the grouping convolution idea of Inception and dilated convolution method are introduced here.

Inception (Szegedy et al., 2015; Szegedy et al., 2017) is a module in Googlenet, as shown in **Figure 4**, which is a locally topologically structured network. Inception performs multiple parallel convolution or pooling operations on the input image and concatenates all the results into a very deep feature map. It uses convolution kernels of different sizes in parallel convolution to obtain different information of the input image, which not only increases the width of the network but also increases the adaptability of the network to scale. The structure of Inception extracts the information of different scales from the input image, enriches the feature information of the image and improves the accuracy of recognition. The structure of Inception has the high-performance characteristics of dense matrix, while maintaining the sparse structure of the network, in order to reduce the computational cost of convolution operation. Therefore, without increasing the complexity of the network, the network can capture more information, retain the original details of more objects, perceive more small-scale feature maps through the perception sparse structure, and improve the recognition accuracy of small object parts while optimizing the neural network.

Dilated Convolution (Yu and Koltun, 2016) increases the reception field by injecting voids into the Convolution map of standard Convolution. Therefore, based on Standard Convolution, Dilated Convolution adds a hyper-parameter called dilation rate, which refers to the number of kernel intervals. The dilated convolution increases the receptive field of the convolution kernel while keeping the number of parameters unchanged so that each convolution output contains a large range of information. At the same time, it can ensure that the size of the output feature map remains unchanged. As shown in the **Figure 5**, the size of the convolution kernel with a dilated rate of 1 remains unchanged, and the receptive field of the 3 × 3 convolution kernel with a dilated rate of 2 is the same as that of the 5 × 5 standard convolution kernel, but the number of parameters is only 9, which is 36% of the number of parameters of the 5 × 5 standard convolution kernel. Compared

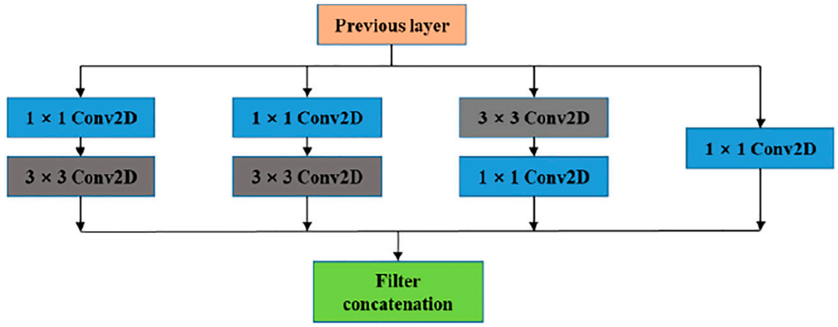


FIGURE 4 | Structure of the Inception module.

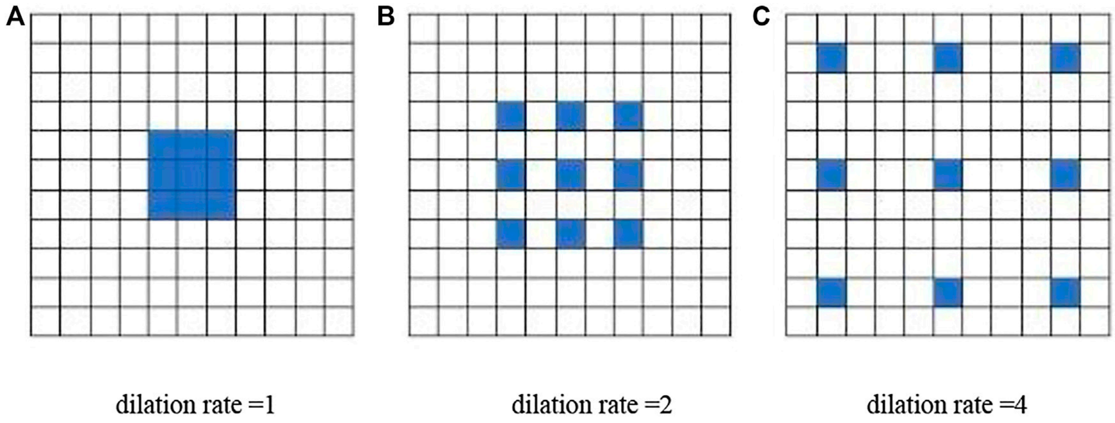


FIGURE 5 | Schematic diagram of dilated convolution with different dilation rates. **(A)** Dilation rate =1 **(B)** dilation rate =2 **(C)** dilation rate =4.

with traditional convolution, dilated convolution can not only preserve the internal structure of data, but also obtain context information, but also will not reduce the spatial resolution (Wang and Ji, 2018). Dilated convolution is also used in WaveNet (van den Oord et al., 2016), bytenet (Kalchbrenner et al., 2016) and other networks to improve network performance.

Using the idea of Inception and dilated convolution, we propose a multi-scale convolution module based on the residual module. As shown in **Figure 6**, the convolution in the module is divided into four groups, and convolution cores of different scales are added. After the image enters the backbone network, features of different scales and depths will be extracted. Compared with the original single convolution method, the possibility of flame features being ignored by the convolution layer is greatly reduced. At the same time, the dilated convolution calculation is added to the standard convolution calculation in the multi-scale convolution module, which can not only simplify the number of weight parameters but also improve the feature extraction ability of the network by improving the receptive field. Dilated convolution will not reduce the spatial resolution. When using multi-size convolution structure, it may affect the resolution and is not conducive to the recognition of small objects. However, if dilated convolution is used, it can effectively avoid the reduction of resolution and strengthen the recognition of small objects. We convert the residual module in the backbone into a multi-scale convolution module, and the rewritten multi-scale convolution module retains the DarknetConv2D and LeakyReLU in the original residual module.

Feature Extraction Structure

Yolov3 extracted three feature layers for object detection, and the output scales of the three feature layers were 52 × 52, 26 × 26, and 13 × 13, respectively. The depth of the corresponding feature layers in the backbone network was located in the middle layer, the middle and lower layer, and the bottom layer. The feature fusion of Yolov3 is a bottom-up one-way path, in this process, the semantic information in the 13 × 13 feature graph is fully utilized after two times of up-sampling and feature fusion. However, for the feature layer with a scale of 52 × 52, it only plays a role in the feature output of its own scale. Therefore, information extraction is missing to some extent. In order to reduce the information missing and ensure the effective extraction of small-scale flame features, the idea of FPN was introduced to improve the feature extraction structure.

FPN(Feature Pyramid Networks) (Lin et al., 2017a) is a Feature extraction structure. As shown in **Figure 7**, FPN carries out multiple feature fusion at different scales. First, feature extraction is carried out from bottom to top, and the

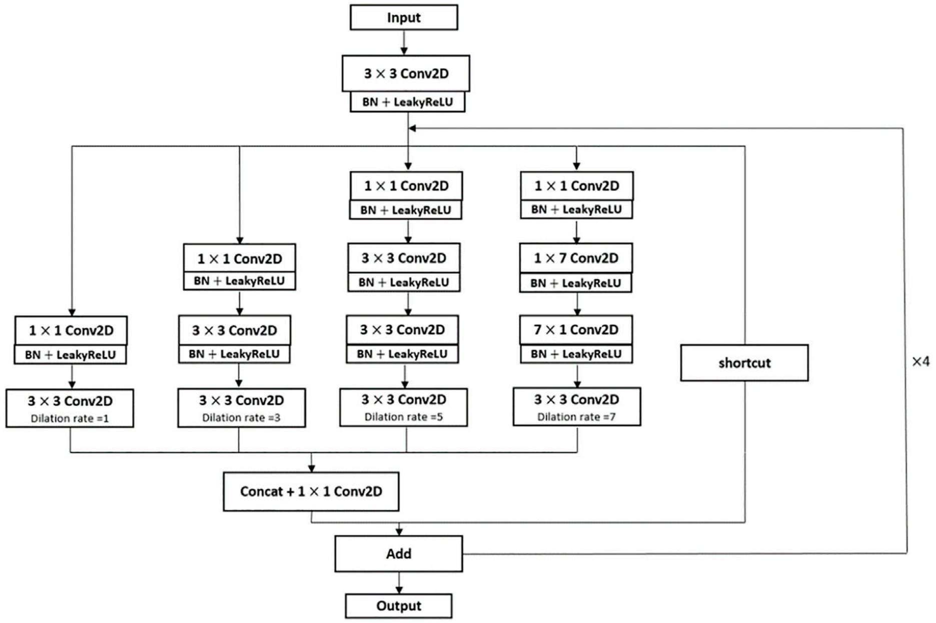


FIGURE 6 | Cyclic structure of multi-scale convolution module.

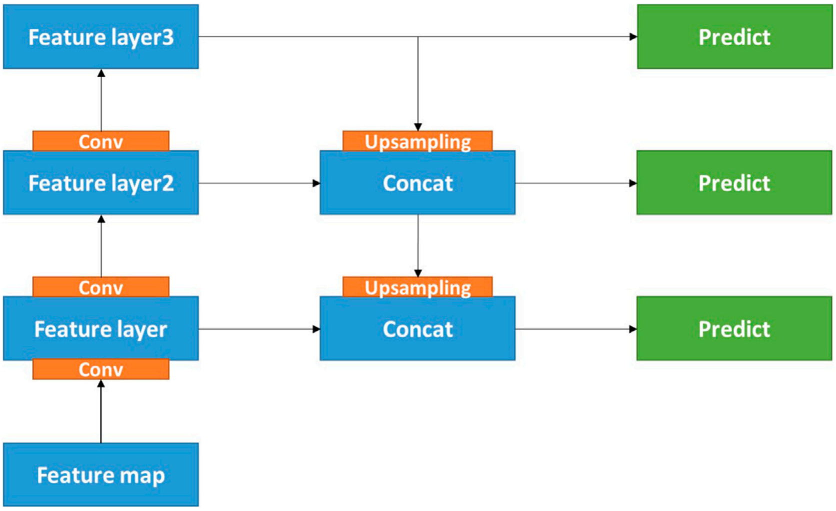


FIGURE 7 | Structure of feature pyramid networks.

scale of the feature map is gradually reduced. After reaching the top level, the feature fusion path is carried out from top to bottom, and the top-level features are up-sampled and gradually merged with the lower level features. It helps to reinforce the low-resolution features of the underlying layer. The idea of feature fusion of FPN has been embodied in many networks.

Inspired by FPN, we expand the one-way feature fusion path in Yolov3 into a two-way feature fusion path. The top-down path is added based on the bottom-up path, which enriches the high-level semantic information and helps detect small flames.

Figure 2 is the structure diagram of the proposed method, in which some residual modules in the backbone network are replaced with multi-scale convolution modules. In terms of feature extraction structure, a bottom-up feature fusion process was carried out first, and a feature map with a scale of 52 × 52 was output. Then, the feature layer was sampled twice and fused with the fusion layer of the 13 × 13 feature layer and the 26 × 26 feature layer, and the result was used as the feature output

TABLE 1 | Flame dataset conditions.

Combustible	Fuel plate size	Interference items	Indoor/outdoor
Polyurethane	—	—	Indoor
Polyurethane	—	Sunlight	Indoor
Polyurethane	—	People	Indoor
Polyurethane	—	Lamplight	Indoor
Polyurethane	—	—	Outdoor
Cardboard	—	—	Indoor
Cardboard	—	Sunlight	Indoor
Cardboard	—	People	Indoor
Cardboard	—	Lamplight	Indoor
Cardboard	—	—	Outdoor
Straw	—	—	Indoor
Straw	—	Sunlight	Indoor
Igniter	—	—	Indoor
Ethanol	7 cm × 7 cm	—	Indoor
Ethanol	7 cm × 7 cm	Sunlight	Indoor
Ethanol	7 cm × 7 cm	People	Indoor
Ethanol	7 cm × 7 cm	Lamplight	Indoor
Ethanol	7 cm × 7 cm	—	Outdoor
Ethanol	15 cm × 15 cm	—	Indoor
Ethanol	15 cm × 15 cm	—	Outdoor
n-heptane	7 cm × 7 cm	—	Indoor
n-heptane	7 cm × 7 cm	Sunlight	Indoor
n-heptane	7 cm × 7 cm	People	Indoor
n-heptane	7 cm × 7 cm	Lamplight	Indoor
n-heptane	7 cm × 7 cm	—	Outdoor
n-heptane	15 cm × 15 cm	—	Indoor
n-heptane	15 cm × 15 cm	—	Outdoor
Toluene	7 cm × 7 cm	—	Indoor
Toluene	7 cm × 7 cm	Sunlight	Indoor
Toluene	7 cm × 7 cm	People	Indoor
Toluene	7 cm × 7 cm	Lamplight	Indoor
toluene	15 cm × 15 cm	—	Indoor

at the 26 × 26 scale. Finally, the output layer is fused with the underlying feature layer as the feature output at the 13 × 13 scale. In this way, the semantic information of the middle layer feature map is enhanced and the model performance is optimized for the detection of small a flame.

EXPERIMENT

Data Set Production

Getting real data sets is not easy for researchers. Currently, the data sets studied are all from several open data sets on the internet. However, there is no standard flame data set for comparison in the field of flame detection (Ghali et al., 2020). Many existing flame data sets on the internet have some problems such as image distortion and excessive flame, which are not conducive to the training of the model and detection of early flame. To better realize the detection of flame by the model and highlight the pertinence of renewable energy fires, we have built a flame data set by ourselves, which includes a variety of combustion conditions under different disturbances in different scenarios.

Flame data sets are mainly divided into two types, indoor and outdoor. In the indoor scene, the standard combustion chamber was selected as the environment for the shooting of the flame video. A standard combustion chamber has a large facility commonly used in the field of fire detection. It is generally used for the research of fuel combustion, fuel products, detectors, and so on. The current utilization forms of renewable energy are mainly renewable energy batteries and new energy vehicles. The battery has a certain fire risk in the process of production, storage and transportation. The warehouse is an important scene in this process. Therefore, for the indoor scene, we make the warehouse scene in the standard combustion room to obtain a similar background. Renewable energy vehicle fire is in a high incidence trend in recent years, so the outdoor scene selects the common parking spots in the campus. Trees, buildings, cars, and other objects are used as the background to obtain the flame data set. Considering the richness of the data set and the robustness of the model, a variety of combustibles and oil plates of different sizes were used to photograph the flames. Sunlight, light, personnel, and other interference items were added into the shooting background, and a variety of combustibles were used to enrich the types of flames. **Table 1** shows the working conditions involved in this data set.

A total of 7,254 images were selected from the filmed videos and used as the training dataset. Some of the images are shown in **Figure 8**. A public LabelImg labeling system was used to label the flame part of the image and store it in the format of Pascal Voc 2007 (Everingham et al., 2006) sample set.

Training

This experiment is carried out under Win10 system, GPU is GeForce GTX 1080, CPU is Intel(R) Core(TM) I7-3960X, 32G memory. Keras, a deep learning framework is used for the model, and Mosaic enhancement is adopted for training. We divided 7,254 data set images into training set and test set, of which 5,558 pictures were used for training and 1,696 pictures were used for testing and verification. The initial learning rate of the training model was set at 0.001.

Evaluation Index

To test the detection performance of the model, we introduced the following indicators: Precision Rate (PR), Recall Rate (RR), Accuracy Rate (AR), and False Alarm Rate (FAR).

The calculation formula of Precision Rate is shown below. In the formula, TP (true positive) refers to the correct response, that is, the number of correctly identified flame pictures in the test set, FP (false positive) refers to the false response, that is, the number of negative samples in the test set that are wrongly identified as flame. In the test, we determine the results of the test set according to the actual situation. We hope that the proposed algorithm can quickly identify all flame objects in the monitoring picture. Therefore, if the intersection of the detection box and the ground truth of the flame object is greater than 0.5 of the union set, it is deemed that the object is correctly detected; otherwise, it misses detection. For a positive sample, we can classify it as *TP* only when all objects in the image are detected. Even if multiple objects are successfully detected, we strictly classify them as *FN* as long as there is one missing object detected because this result does not meet our requirements. For a negative sample, if there is no detection box, it can be

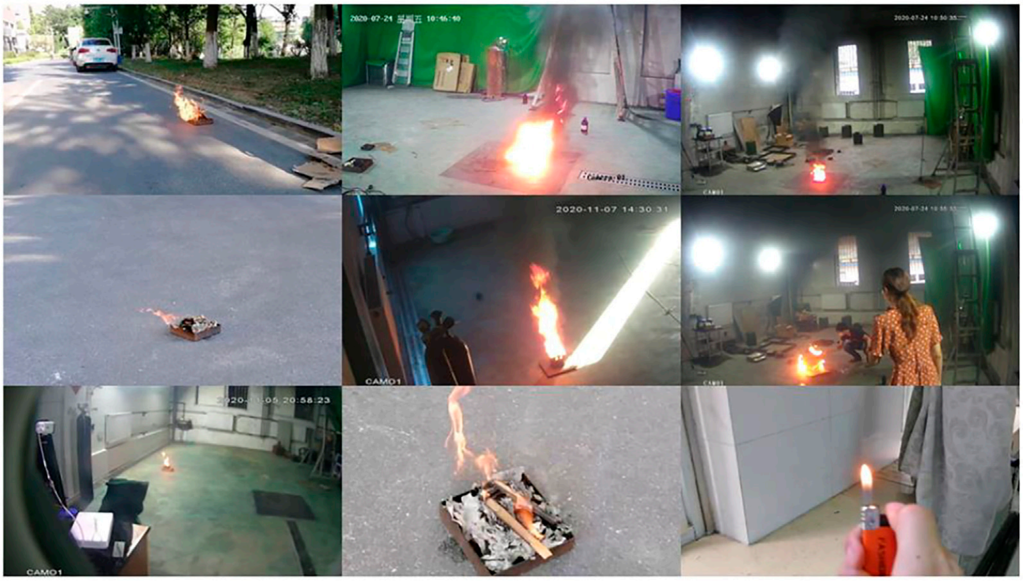


FIGURE 8 | Some images of the flame dataset.

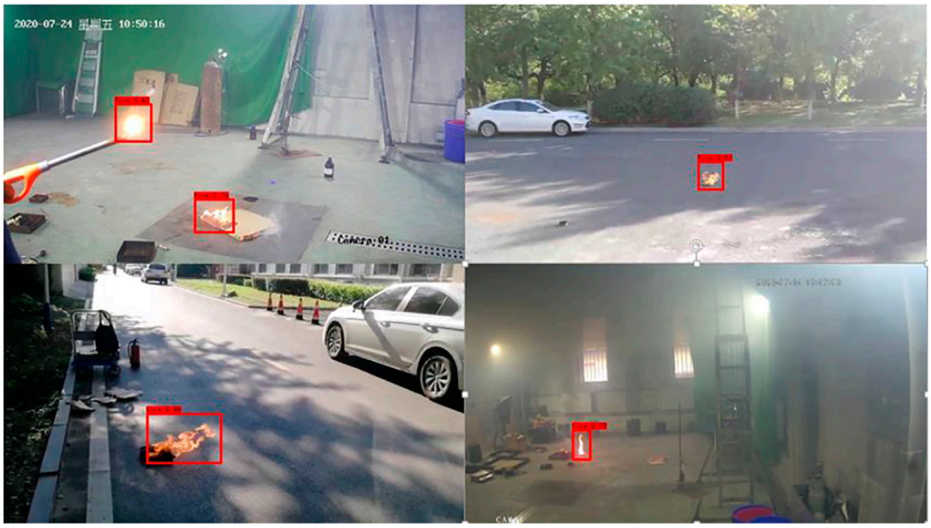


FIGURE 9 | Partial results identified by the proposed model.

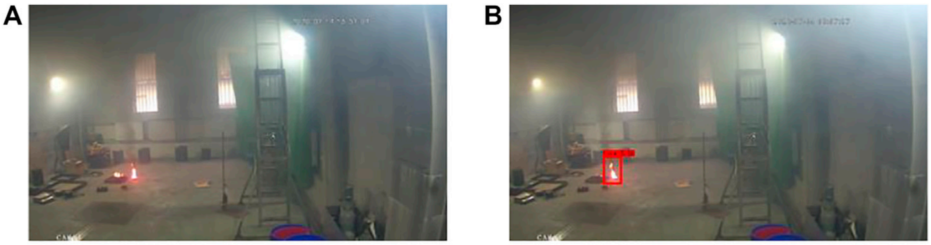
classified as *TN*, and if there is a detection box, it can be classified as *FP*. Precision Rate represents the proportion of correctly detected flame in all detection results, reflecting the credibility of flame detection.

$$\mathrm{PR} = TP/(TP + FP) \times 100\%$$

The formula for Recall Rate is defined as follows. In the formula, FN (false negative) refers to the wrong negative sample, that is, the number of flame images that are not recognized in the test set. Recall Rate represents the proportion of correctly detected flames in all fires that should have been detected, reflecting the model's ability to detect flames.

$$\mathrm{RR} = TP/(TP + FN) \times 100\%$$

The calculation formula for Accuracy Rate is shown below, in which TN (true negative) refers to the correct negative sample, that is, the number of negative samples without false positives. Accuracy Rate refers to the ratio of correctly predicted samples to the total predicted samples, which reflects the comprehensive ability of model detection.



the image frame identified as a negative sample in other models

the image frame identified as a positive sample in the proposed model

FIGURE 10 | Recognition results of small flame image frames in different models. **(A)** The image frame identified as a negative sample in other models **(B)** the image frame identified as a positive sample in the proposed model.

TABLE 2 | Test results.

Model	PR (%)	RR (%)	FAR(%)	AR (%)
Proposed method	98.7	93.7	1.2	96.3
Yolov3 (Redmon and Farhadi, 2018)	93.4	92.5	6.2	93.2
SSD (Liu et al., 2016)	94.5	51.9	2.8	74.9
RFBnet (Liu et al., 2018)	90.4	86.1	8.8	88.7
Efficientdet (Tan et al., 2019)	95.9	93.3	3.8	94.8
Yolov4 (Bochkovskiy et al., 2004)	93.1	92.7	6.6	87.1
Retinanet (Lin et al., 2017b)	96.5	88.7	3.1	92.9
Faster R-CNN (Ren et al., 2016)	97.3	93.1	2.3	95.2
Fast R-CNN (Girshick, 2015)	94.3	85.2	3.5	90.1

$$\mathrm{AR} = (TP + TN)/(TP + TN + FP + FN) \times 100\%$$

The formula for False Alarm Rate is defined as follows. False Alarm Rate is an evaluation index in the field of fire detection. For the application scenarios of fire detection, most of the time is in the non-flame negative sample state, so it is very important to control false positives for fire detection.

$$\mathrm{FAR} = FP/(FP + TN) \times 100\%$$

Test With Test Set

The trained model was used to test the test set, and part of the test results were shown in **Figure 9**. The proposed method works well in different scenarios, it can be seen that the model has high accuracy in the identification and location of small flames, which is helpful to detect and alarm in the early stage of fire.

To better evaluate the performance of the proposed model, in addition to Yolov3, other common one-stage target detection models are introduced for comparative testing. The same training set was used to train these models in the same environment, and the same test set was used for testing. The results show that the proposed method is more sensitive to small flames. As shown in **Figure 10**, other models cannot successfully identify such image frames with a small fire, but in our proposed model, they can be successfully identified as fire. More specific results are shown in **Table 2**. Both the training set and the test set are small flame images. It can be seen that compared with Yolov3, the proposed method has improved in all four indicators, including a significant increase in Precision Rate and a significant decrease in False Alarm Rate, reflecting that the improved model has indeed improved the performance of detecting small flames. Compared with other models, the proposed method also shows some advantages, with all four indexes ranking first, reflecting the absolute superiority of our method in early flame detection. Precision Rate and False Alarm Rate were superior, with the false alarm rate as low as 1.2%, indicating the stability of the method. Besides, we add two advanced two-stage models as a comparison. The two-stage model first generates a series of candidate boxes as samples by the algorithm, and then classifies samples by convolution neural network. Therefore, it has higher accuracy and slower speed. It can be seen that the proposed method has higher accuracy while having smaller size and faster calculation speed.

Real-Time Test of Fire Scenarios

In order to test the effect of the model we proposed in practical application, we used the monitoring cameras installed in the laboratory building to carry out real-time flame detection, which were similarly divided into an indoor scene and an outdoor scene.

As shown in the **Figure 11**, the outdoor scenes include the rooftop and the outdoor scene of the first floor. The rooftop is equipped with three surveillance cameras at different angles, while the outdoor scene of the first floor is equipped with one surveillance camera. The interior scene is a standard combustion chamber with a surveillance camera installed inside. Since the test object is small flame, we choose an oil pan with the size of 7 cm × 7 cm to ignite n-heptane for the test.

We tested the non-fire scenario in each environment before ignition, and no false positives were generated. So, PR and FAR were not included in the analysis of the results, only the recall rate (RR) was analyzed. The total number of frames in the real-time detection process and the number of fire frames detected were calculated by the script, and the recall rate was calculated accordingly. In real-time detection, we introduce the concept of FPS, FPS is the number of image frames detected per second, reflecting the speed of model detection. The test results are shown in **Table 3**. The model performs well in real-time detection. As shown in **Figure 12**, the majority of small flames can be successfully identified in real-time detection, and the FPS

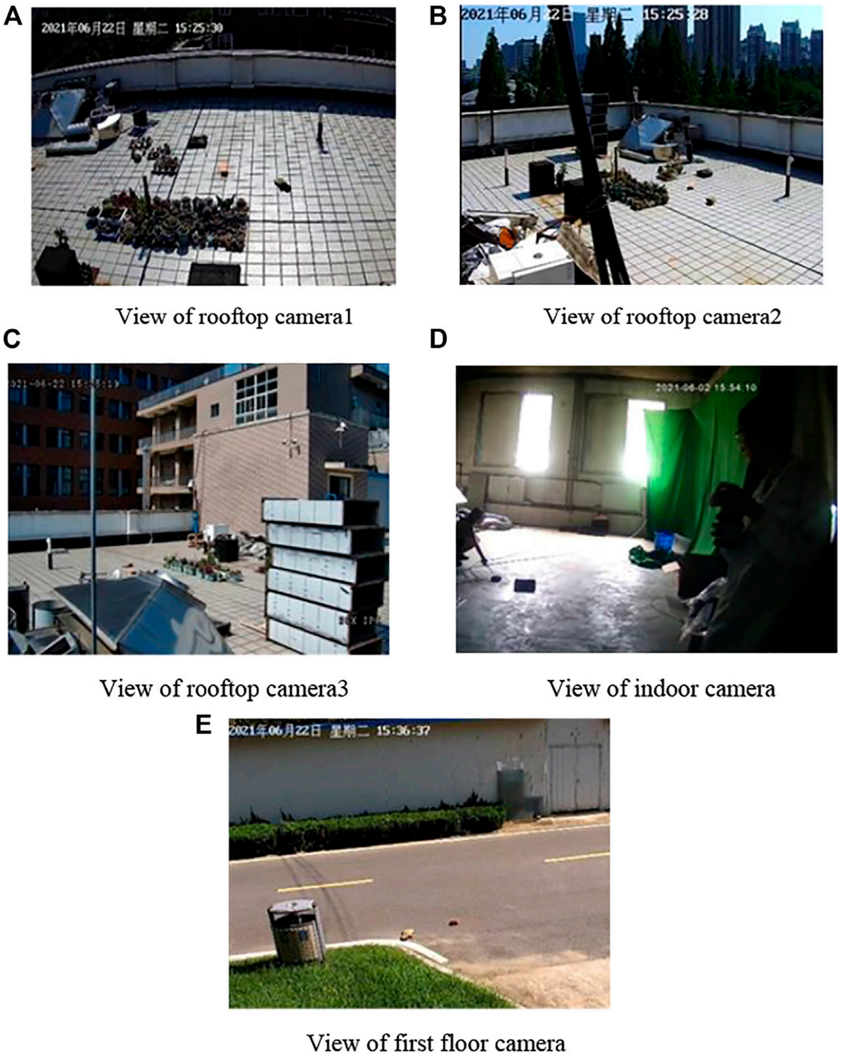

A
2021年06月22日 星期二 15:25:30
View of rooftop camera1
B
2021年06月22日 星期二 15:25:28
View of rooftop camera2
C
View of rooftop camera3
D
2021-06-02 15:54:10
View of indoor camera
E
2021年06月22日 星期二 15:36:37
View of first floor camera

FIGURE 11 | Views of monitoring cameras. **(A)** View of rooftop camera1 **(B)** view of rooftop camera2 **(C)** View of rooftop camera3 **(D)** view of indoor camera **(E)** view of first floor camera.

TABLE 3 | Real-time detection results.

Camera	Fuel	RR (%)
Indoor camera	n-heptane	78.4
Rooftop camera1	n-heptane	70.2
Rooftop camera2	n-heptane	65.6
Rooftop camera3	n-heptane	66.3
First floor camera	n-heptane	76.3

value is stable at around 11, which reflects the sensitivity of the proposed model to small flames. However, there is still a gap between the recall rate in real-time detection and the recall rate in the test set.

Therefore, we analyzed the flame frames that were not detected. As shown in **Figure 13**, undetected flame frames fall into two categories. One is that it is difficult to identify the flame due to the complex background. This situation mainly occurs in two of the monitoring pictures on the rooftop. The monitoring perspective of these two cameras leads to a chaotic picture, which also explains why the real-time detection performance is better in the standard room with a relatively empty environment. In this case, we can consider labeling the undetected flame frames and iterating training. One is due to the influence of light, the flame is blurred or even the flame becomes invisible, so they cannot be detected. For flame frames whose flame profile is still recognizable under the influence of light, we can label them and then carry out iterative training of the model. As for the flame frame that is not visible under the influence of light, this problem is difficult to be solved in a single visible light channel. In the future, we will seek a solution by combining infrared channels and visible light channels.

In the training process of neural network, iteration refers to the process of updating the parameters of the model with a batch

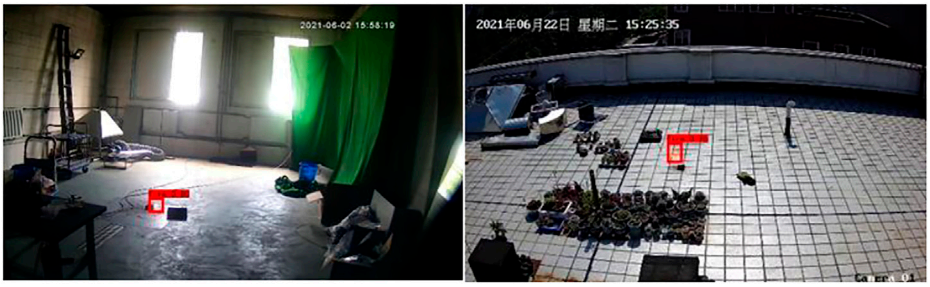
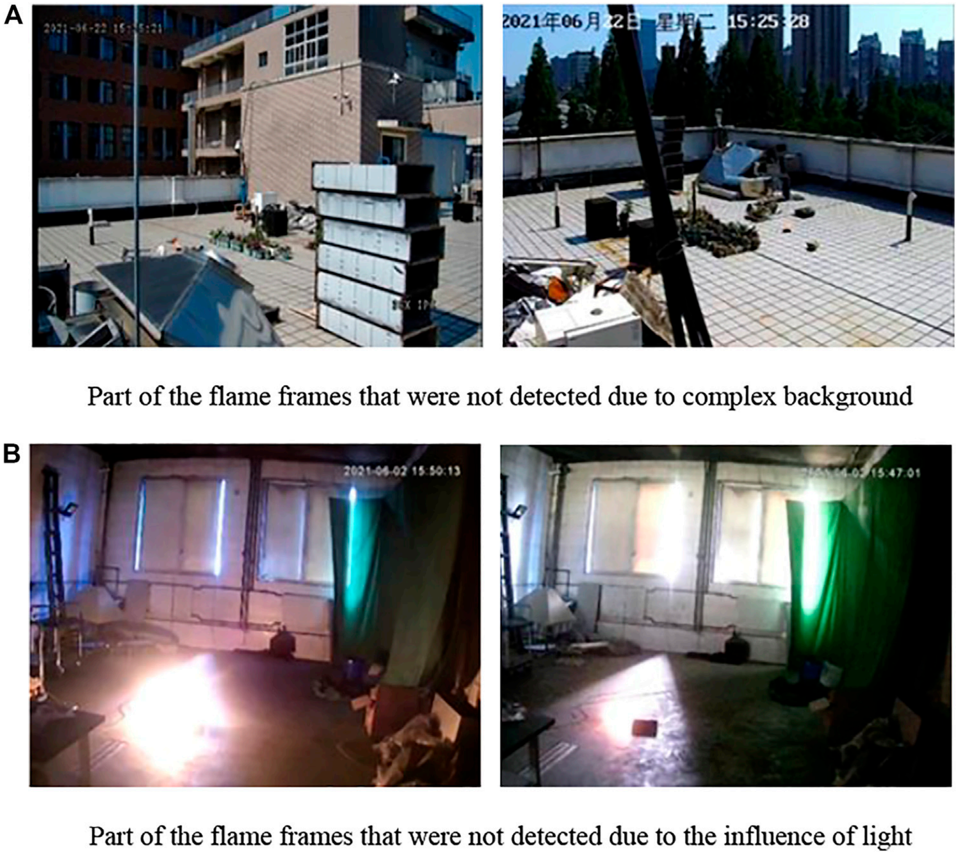


FIGURE 12 | Part of the flame frames that were detected.



Part of the flame frames that were not detected due to complex background

Part of the flame frames that were not detected due to the influence of light

FIGURE 13 | Part of the flame frames that were not detected. **(A)** Part of the flame frames that were not detected due to complex background **(B)** part of the flame frames that were not detected due to the influence of light.

TABLE 4 | Real-time detection results (after iterative training).

Camera	Fuel	RR (%)
Indoor camera	n-heptane	91.3
Rooftop camera1	n-heptane	89.1
Rooftop camera2	n-heptane	88.7
Rooftop camera3	n-heptane	88.3
First floor camera	n-heptane	90.5

of data. In practical engineering applications, because the detection environment is in a stable state for a long time, we collect the missing flame frames in the detection and update the existing data set. On this basis, the training is continued with the new data set to update the model parameters, so that the model can recognize the unrecognized flame frames and achieve better actual detection effect in this environment. We call this training method iterative training. After the iterative training, the detection rate of the flame frame has been significantly improved when the same scene is detected in real-time. The results are shown in **Table 4**. The recall rate has now approached the value in the test set. This shows that in practical applications, model iteration with pictures of actual scenes can achieve the best detection performance of the model.

CONCLUSION

In order to solve the problem that small flames in early fires are prone to omission and false positives, this paper proposes an improved model based on Yolov3 for this problem. Multi-scale convolution and increasing receptive field were used to improve the sensitivity of the model to a small flame, and FPN structure was used to enhance the ability of feature extraction. The experimental results show that both compared with the original Yolov3 model and other commonly used object detection models, the proposed model performs better in flame recognition and accomplishes its original design intention for small flame recognition. In this paper, the proposed model is applied to the actual scene to obtain good performance, found iterative training for practical application has a key role in testing. At the same time, this paper also establishes a flame data set for early fires, including indoor and outdoor conditions, which provides a certain basis for future flame detection research.

In the process of establishing the data set, we also found the deficiency of the current model. In the case of direct interference from some strong light sources, the flame and light are fused and cannot be distinguished from the naked eye, which cannot be started from the annotation of the data set. How to solve such problems will be the research target of the next stage.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

QZ and PD conceived of the presented idea. PD. and YH developed the theory and performed the experiments. GL. and RT. verified the analytical methods. YZ encouraged PD and MS to investigate real-time detection and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

FUNDING

This work was financially supported by the National Key Research and Development Plan under Grant No. 2018YFC0809502, the Open Project Program of State Key Laboratory of Fire Science under Grant No. HZ2020-KF07, and the Research Plan of Fire and Rescue Department, Ministry of Emergency Management under Grant No. 2018XFGG12. National Natural Science Foundation of China (No. 52076084).

REFERENCES

Abohamzeh, E., Salehi, F., and Sheikholeslami, M. (2021). Review of Hydrogen Safety during Storage, Transmission, and Applications Processes. *J. Loss Prev. Process Industries* 72, 72. doi:10.1016/j.jlp.2021.104569

Bochkovskiy, A., Wang, C-Y., and Liao, H-Y. M. (2004). *YOLOv4: Optimal Speed and Accuracy of Object Detection*. arxiv is 2004.10934. Available at: https://ui.adsabs.harvard.edu/abs/2020arXiv200410934B

Borges, P. V. K., and Izquierdo, E. (2010). A Probabilistic Approach for Vision-Based Fire Detection in Videos. *IEEE Trans. Circuits Syst. Video Technol.* 20, 721–731. doi:10.1109/tcsvt.2010.2045813

Dai, J., Li, Y., and He, K. (2016). “R-FCN: Object Detection *via* Region-Based Fully Convolutional Networks,” in Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 2016 (Barcelona: Spain: Curran Associates Inc), 379.

Dimitropoulos, K., Barmpoutis, P., and Grammalidis, N. (2014). Spatio-Temporal Flame Modeling and Dynamic Texture Analysis for Automatic Video-Based Fire Detection. *IEEE Trans. Circuits Syst. Video Techn.* 2015 (25), 339–351. doi:10.1109/TCSVT

Dua, M., Kumar, M., and Charan, G. S. (2020). “An Improved Approach for Fire Detection Using Deep Learning Models,” in 2020 International Conference on Industry 40 Technology (I4Tech), Pune, India, February 13-15, 2020.

Everingham, M., Zisserman, A., and Williams, C. K. I. (2006). “The 2005 PASCAL Visual Object Classes challenge,” in *Machine Learning Challenges—Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Textual Entailment* (Berlin: Springer).

Fang, J., Cai, J. N., and He, X. Z. (2021). Experimental Study on the Vertical thermal Runaway Propagation in Cylindrical Lithium-Ion Batteries: Effects of Spacing and State of Charge. *Appl. Therm. Eng.* 197, 197. doi:10.1016/j.applthermaleng.2021.117399

Frizzi, S., Kaabi, R., and Bouchouicha, M. (2016). “Convolutional Neural Network for Video Fire and Smoke Detection,” in IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, October 23-26, 2016, 877–882. doi:10.1109/iecon.2016.7793196

Ghali, R., Jmal, M., and Souidene Mseddi, W. (2020). *Recent Advances in Fire Detection and Monitoring Systems: A Review. in: Cham*. Springer International Publishing.

Girshick, R. (2015). *Fast R-CNN*. arxiv is 1504.08083. Available at: https://ui.adsabs.harvard.edu/abs/2015arXiv150408083G.

Kalchbrenner, N., Espeholt, L., and Simonyan, K. (2016). *Neural Machine Translation in Linear Time*. arxiv is 1610.10099. Available at: https://ui.adsabs.harvard.edu/abs/2016arXiv161010099K

Khudayberdiev, O., and Butt, M. H. F. (2020). Fire Detection in Surveillance Videos Using a Combination with PCA and CNN. *Acad. J. Comput. Inf. Sci.* 3, 3. doi:10.25236/AJCIS.030304

Kim, B., and Lee, J. A. (2019). *Video-Based Fire Detection Using Deep Learning Models*. Basel, Switzerland: Applied Sciences.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based Learning Applied to Document Recognition. *Proc. IEEE* 86, 2278–2324. doi:10.1109/5.726791

Li, P., and Zhao, W. (2020). Image Fire Detection Algorithms Based on Convolutional Neural Networks. *Case Studies in Thermal Engineering*. Amsterdam, Netherlands: Elsevier, 19. doi:10.1016/j.csite.2020.100625

Lin, G., Zhang, Y., Xu, G., and Zhang, Q. (2019). Smoke Detection on Video Sequences Using 3D Convolutional Neural Networks. *Fire Technol.* 55, 1827–1847. doi:10.1007/s10694-019-00832-w

Lin, T., Dollár, P., and Girshick, R. (2017). “Focal Loss for Dense Object Detection,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 21-26, Venice, Italy, October 22-29, 2017.

Lin, T., Goyal, P., and Girshick, R. Feature Pyramid Networks for Object Detection." in 2017 IEEE International Conference on Computer Vision (ICCV) 22-29, Honolulu, HI, July 21-26, 2017.

Liu, S., Huang, D., and Wang, Y. (2018). "Receptive Field Block Net for Accurate and Fast Object Detection," in *Cham* (Springer International Publishing) doi:10.1007/978-3-030-01252-6_24

Liu, W., Anguelov, D., and Erhan, D. (2016). "SSD: Single Shot MultiBox Detector," in *Cham* (Springer International Publishing), doi:10.1007/978-3-319-46448-0_2

Ould Ely, T., Kamzabek, D., and Chakraborty, D. (2019). Batteries Safety: Recent Progress and Current Challenges. *Front. Energ. Res.* 7, 7. doi:10.3389/fenrg.2019.00071

Qazi, A., Hussain, F., Rahim, N. A., Hardaker, G., Alghazzawi, D., Shaban, K., et al. (2019). Towards Sustainable Energy: A Systematic Review of Renewable Energy Sources, Technologies, and Public Opinions. *Ieee Access* 7, 63837–63851. doi:10.1109/access.2019.2906402

Redmon, J., and Farhadi, A. (2018). *YOLOv3: An Incremental Improvement*. arXiv: 1804.02767. Available at: https://ui.adsabs.harvard.edu/abs/2018arXiv180402767R

Ren, S., He, K., Girshick, R., and Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach Intell.* 39 (39), 1137–1149. doi:10.1109/TPAMI10.1109/TPAMI.2016.2577031

Sharma, J., Granmo, O-C., and Goodwin, M. (2017). *Deep Convolutional Neural Networks for Fire Detection in Images*. Springer International Publishing

Shen, D., Chen, X., and Nguyen, M. (2018). "Flame Detection Using Deep Learning," in 2018 4th International Conference on Control, Automation and Robotics (ICCAR) 20-23, Auckland, New zealand, April 20-23, 2018.

Szegedy, C., Ioffe, S., and Vanhoucke, V. (2017). *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. San Francisco, CaliforniaUSA: AAAI Press.

Szegedy, C., Wei, L., and Yangqing, J. (2015). "Going Deeper with Convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 7-12, San Francisco, CA, February 04-09, 2017.

Tan, M., Pang, R., and Le, Q. V. (2019). *EfficientDet: Scalable and Efficient Object Detection*. arXiv: 191109070T. Available at: https://ui.adsabs.harvard.edu/abs/2019arXiv191109070T

Tran, D., Bourdev, L., and Fergus, R. (2015). "Learning Spatiotemporal Features with 3D Convolutional Networks," in 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, December, 11-18, 2015.

van den Oord, A., Dieleman, S., and Zen, H. (2016). *WaveNet: A Generative Model for Raw Audio*. arXiv :1609.03499. Available at: https://ui.adsabs.harvard.edu/abs/2016arXiv160903499V

Wang, Z. Y., and Ji, S. W. (2018). "Smoothed Dilated Convolutions for Improved Dense Prediction," in 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) Londonproceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining, London, England, August 19-23, 1999 (ENGLAND).

Xu, G. (2020). *Research on Deep Domain Adaptation and Saliency Detection in Fire Smoke Image Recognition*. University of Science and Technology of China.

Yamagishi, H., and Yamaguchi, J. (1999). "Fire Flame Detection Algorithm Using a Color Camera," in MHS'99 Proceedings of 1999 International Symposium on Micromechatronics and Human Science (Cat No99TH8478), Nagoya, Japan, November 23-26, 1999.

Yang, H., Leow, W. R., and Chen, X. (2018). Thermal-responsive Polymers for Enhancing Safety of Electrochemical Storage Devices. *Adv. Mater.* 30, e1704347. doi:10.1002/adma.201704347

Young-Jin, K., and Eun-Gyung, K. (2017). *Fire Detection System Using Faster R-CNN*. Seoul: Korea Information and Communication Society.

Yu, F., and Koltun, V. (2016). *Multi-Scale Context Aggregation by Dilated Convolutions*. arxiv is 1511.07122. Available at: https://ui.adsabs.harvard.edu/abs/2015arXiv151107122Y

Zhang, X., Gao, Y., and Wang, H. (2020). Improve YOLOv3 Using Dilated Spatial Pyramid Module for Multi-Scale Object Detection. *Int. J. Adv. Robotic Syst.* 17, 1729881420936062. doi:10.1177/1729881420936062

Zhong, C., Chen, Z., and Yu, S. (2020). Video Fire Recognition Based on Multi-Channel Convolutional Neural Network. *J. Phys. Conf. Ser.* 1634, 1634. doi:10.1088/1742-6596/1634/1/012020

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.